Advanced Engineering Informatics 74 (2026) 104635

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

LDS-former: A lightweight dual-stream transformer for real-time acoustic emission monitoring of crack evolution in offshore steel structures

Yuchen Lu [a,1], Zhenhao Zhu [a,1], Yifei Li [b], Hongbing Liu [a,*], Jiufan Hou [a], Chunyu Zhou [a], M.Abdel Wahab [c,d]

[a] Yantai Research Institute, Harbin Engineering University, Yantai 264005, China
[b] School of Engineering, Huzhou University, Huzhou 313000, China
[c] Soete Laboratory, Department of Electrical Energy, Metals, Mechanical Constructions, and Systems, Faculty of Engineering and Architecture, Ghent University, Gent, Belgium
[d] Parallel and Distributed Systems Laboratory, Jožef Stefan Institute, Jamova cesta 39, Ljubljana SI-1000, Slovenia

ARTICLE INFO

Keywords:
Structure health monitoring
Acoustic emission
Fatigue crack detection
Lightweight transformer

ABSTRACT

Fatigue damage under cyclic loading remains a critical failure driver in offshore energy infrastructures, where dynamic stresses accelerate crack initiation and propagation, necessitating real-time diagnostic technologies to ensure service safety while limited computational capacity at edge nodes constrains monitoring system deployment. Acoustic emission (AE) technology achieves damage monitoring by capturing stress wave release during material fracture processes, yet AE signals exhibit transient-evolutionary duality where high-frequency bursts encode instantaneous fracture events while low-frequency trends reflect gradual damage accumulation, demanding intelligent diagnostic models to simultaneously capture heterogeneous temporal patterns while existing deep learning methods struggle to jointly resolve this dual characteristic and their computational demands prohibit edge deployment. To address this challenge, this study introduces a mechanism-aligned lightweight dual-stream transformer (LDS-Former) that matches the physical duality of AE signals through dual-stream architecture, where the local stream preserves transient waveform details while the global stream models long-range dependencies, with a gated fusion mechanism adaptively weighting contributions to optimize feature representation. LDS-Former achieves 95.05% accuracy with only 0.199 M parameters and 5.215 M FLOPs, accomplishing 97.2% parameter reduction relative to deep architectures and 6.2% accuracy improvement over state-of-the-art lightweight methods. LDS-Former retains 86.24% accuracy with only 20% training data and maintains over 84% accuracy under extreme noise conditions. Edge deployment validation shows that LDS-Former achieves 2.086 ms inference latency on the STM32H7B3I-DK microcontroller with flash and RAM occupation of only 845.0 KB and 100.3 KB respectively. These findings confirm that LDS-Former, through synergistic integration of signal physical mechanisms and deep learning, provides an intelligent solution combining diagnostic precision with computational efficiency for offshore structural health monitoring, advancing practical deployment of edge intelligence in industrial safety monitoring.

1. Introduction

Steel structures serve as critical load-bearing components in offshore energy infrastructures, where structural joints experience fatigue damage accumulation under cyclic loading, leading to fatigue failure that exhibits both sudden onset and irreversible consequences[1,2]. Subjected to hostile marine environments, platform structures endure persistent cyclic loading and stochastic wave forces[3,4]. Under such dynamic loading conditions, microscopic defects in structures undergo continuous accumulation, ultimately leading to cumulative fatigue damage. Once cracks undergo unstable propagation, catastrophic failure poses severe risks to personnel safety and environmental integrity[5,6]. Therefore, employing non-destructive testing techniques to establish structural health monitoring (SHM) systems for early warning of fatigue damage has become a technological imperative for ensuring offshore platform service safety[7,8]. Acoustic emission (AE) technology has

* Corresponding author.
E-mail addresses: luyuchen6@foxmail.com (Y. Lu), hb_liu@hrbeu.edu.cn (H. Liu).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.aei.2026.104635
Received 4 December 2025; Received in revised form 12 February 2026; Accepted 24 March 2026
Available online 31 March 2026
1474-0346/

emerged as a core method for fatigue damage monitoring through its inherent capability to capture transient stress wave release during material fracture processes[9–12].

In recent years, deep learning methods have achieved significant progress in intelligent recognition of acoustic emission signals[13–18]. For example, CNNs have been successfully applied to recognition of AE signal waveform details through their local feature extraction capability [19,20], while Transformer architectures have achieved effective capture of long-sequence global dependencies in AE signals through self-attention mechanisms[21,22]. However, existing methods face fundamental challenges when addressing steel structural crack identification tasks. The root cause lies in the transient-evolutionary duality of AE signals, which contain both high-frequency transient pulses characterizing instantaneous crack fracture at contact interfaces and low-frequency evolutionary trends reflecting friction-driven damage accumulation processes, with these two feature types spanning vastly different time scales and frequency ranges[23–25]. This duality introduces heterogeneous feature requirements that demand simultaneous capture of local discriminability and global dependency modeling. Traditional CNNs struggle to capture long-range dependencies in crack evolution due to restricted receptive fields imposed by local convolution kernels[26,27], while pure attention models, despite excelling at global dependency capture, fail to preserve microscopic waveform details owing to the high-frequency information loss inherent in tokenization processes[28]. Consequently, existing single-stream architectures fail to reconcile these heterogeneous requirements, unable to simultaneously balance local discriminability with global dependency capture, leading to performance stagnation in steel structural crack identification.

Beyond feature extraction challenges, computational constraints in edge computing further exacerbate model deployment difficulties. The imperative for high sampling rates to avoid missing minor crack propagation signals inevitably results in lengthy temporal sequences[29–31], which incur quadratic complexity in Transformer self-attention mechanisms, imposing prohibitive computational burden[32]. However, offshore platform monitoring relies on resource-constrained edge nodes constrained by strict computational budgets[33–38]. While existing high-precision models excel in laboratory settings, their prohibitive computational costs render them unsuitable for online real-time monitoring[39–42]. This accuracy-efficiency trade-off between theoretical performance and practical deployment hinders the practical adoption of deep learning techniques in marine engineering.

To address these two major technical challenges, this paper proposes the Lightweight Dual-Stream Transformer, termed LDS-Former. Guided by signal physics, this method employs a mechanism-aligned lightweight dual-stream network architecture specifically designed for AE signal identification of steel structural crack propagation. LDS-Former adopts a dual-pathway design with local stream and global stream. To ensure discriminability of AE signals, the local stream captures waveform details of transient pulses through inverted residual convolutions, while the global stream captures long-range dependencies of evolutionary trends through attention mechanisms. The two pathways achieve adaptive fusion through a gating mechanism, thereby realizing synergistic perception of microscopic and macroscopic features in AE signals. To ensure computational efficiency, we design a manifold projection mechanism incorporating Stem modules and sequence compression to map lengthy temporal sequences into compact feature spaces, substantially reducing computational burden while preserving discriminative semantics of AE signals and breaking through edge-side computational constraints. Furthermore, addressing the engineering reality that real offshore platforms involve diverse component geometries, this paper constructs an AE-based physical fatigue experimental dataset encompassing diverse component geometries and complete damage lifecycles, systematically validating the industrial robustness of the method under cross-component generalization and class imbalance scenarios. The main contributions of this paper include:

- Proposing the LDS-Former network architecture specifically designed for steel structural crack propagation identification, which realizes synergistic perception of microscopic waveform and macroscopic evolutionary features through a mechanism-aligned dual-stream design, resolving the challenge of simultaneously accommodating dual-scale characteristics in AE signals.
- Developing sequence compression and manifold projection mechanisms for high-sampling-rate AE signals, which substantially reduce computational complexity while preserving discriminative semantics, overcoming computational constraints in deploying long-sequence monitoring models on resource-constrained edge nodes.
- Constructing an AE-based physical fatigue experimental dataset encompassing diverse component geometries and complete damage lifecycles, validating the industrial robustness of the model under cross-component generalization and class imbalance scenarios.

2. Proposed method

2.1. Problem formulation and overall framework

Steel structural crack propagation intelligent recognition is formulated as mapping learning from temporal signal space $\mathscr{X} = \mathbb{R}^T$ to damage state space $\mathscr{Y} = \{1, 2, \ldots, K\}$. The variable T denotes sampling sequence length and K denotes number of damage categories encompassing pre-fracture states and fracture states across component types and crack configurations. Given training set $\mathscr{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$, the objective is to learn discriminative function $f : \mathscr{X} \rightarrow \mathscr{Y}$ minimizing empirical risk $\mathscr{L}(f) = \frac{1}{N}\sum_{i=1}^N \ell(f(\boldsymbol{x}_i), y_i)$.

Steel structural fracture acoustic emission signals $\boldsymbol{x}(t) \in \mathbb{R}^T$ originate from transient stress wave release and multi-path propagation within materials, thereby exhibiting transient-evolutionary duality where instantaneous crack propagation generates high-frequency transient pulses whereas fatigue evolution manifests as low-frequency envelope modulation. This duality is characterized by multi-scale coupling where symmetric stress fields excite single dominant frequencies whereas asymmetric stress gradients and complex cross-sectional topologies induce coexistence of multi-scale frequency components. Consequently, steel structural crack propagation intelligent recognition models must simultaneously achieve local discriminability and global dependency modeling. However, lightweight convolutional networks are inherently constrained by local receptive fields whereas pure attention models fail to preserve transient details. Existing methods cannot balance dual requirements under lightweight constraints.

To address these challenges, we propose LDS-Former model satisfying dual modeling requirements through synergistic perception of local and global features. LDS-Former adopts a hierarchical dual-stream architecture comprising four functional modules in coordination. First, the Stem module embeds single-channel signals into multi-channel feature representations. Subsequently, the three-stage backbone extracts multi-scale features through progressive downsampling. Within each stage, LDS-Former blocks decouple local stream and global stream then fuse them via gated mechanism. Finally, the classification head aggregates features to generate decisions. Specifically, the local stream preserves transient pulses through inverted residual convolutions, the global stream captures long-range dependencies through compressed-sequence attention, whereas convolution-enhanced mechanism preserves temporal structure within the global stream. Fig. 1 illustrates the hierarchical architecture of LDS-Former.

2.2. Frequency-domain rationale for dual-stream design

Transient pulses and evolutionary trends in acoustic emission signals manifest fundamental differences in the time–frequency domain. Transient pulses correspond to high-frequency energy bursts from instantaneous crack fracture. Their waveform characteristics concentrate within

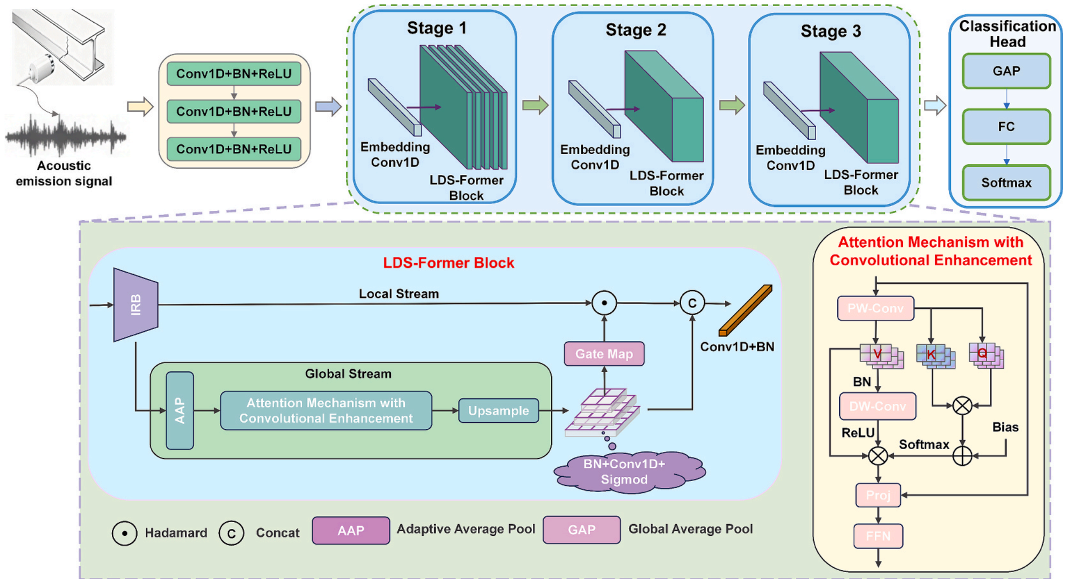


Fig. 1. Overall architecture of the proposed LDS-Former network for acoustic emission signal classification.

extremely short time windows. Capturing such rapid amplitude variations requires precise temporal localization capability. Evolution trends reflect the gradual process of damage accumulation. They manifest as slow modulation of low-frequency envelopes. Discriminative information distributes over extended time spans. Identifying evolution patterns relies on fine resolution of frequency components. These two feature types pose contradictory requirements on analysis window selection. Short time windows can lock transient event occurrences but lack frequency resolution capability. Long time windows can precisely distinguish frequency components yet smooth out transient details. The frequency-domain characteristics of convolutional operators directly stem from the compactness of their temporal kernel functions. Convolution with kernel width τ produces frequency-domain response with bandwidth $\Delta f \propto 1/\tau$. The local stream employs small-scale convolutions with kernel width $\tau = 3$. This narrow-kernel design corresponds to wide-band response. The filter maintains sufficient passband gain across high-frequency intervals. This avoids the low-pass smoothing effect introduced by traditional long-kernel convolutions. The wide-band characteristic ensures that high-frequency components of transient pulses are preserved during feature extraction. Attention mechanisms essentially execute global weighted averaging over sequences $\boldsymbol{y}_t = \sum_{s=1}^{L} \alpha_{ts} \boldsymbol{x}_s$. The normalization condition $\sum_s \alpha_{ts} = 1$ renders it equivalent to an adaptive weighted filter. When weights α_{ts} exhibit smooth distribution patterns over long sequences the weighted summation operation naturally suppresses high-frequency oscillatory components in sequences while amplifying low-frequency trend terms. The global stream compresses sequence length from L to L_g through adaptive pooling. This downsampling operation reduces the effective sampling rate. The representable maximum frequency shrinks proportionally by L_g/L. This focuses computational resources on low-frequency evolution patterns. The dual-stream design achieves functional specialization through separated operations. The local stream leverages the wide-band characteristic of short-kernel convolutions to preserve transient pulses. The global stream leverages the combination of long-sequence attention and sequence compression to enhance low-frequency trends. Their complementary response characteristics in the frequency domain align with the physical duality of acoustic emission signals.

2.3. Stem module for initial feature extraction

Raw steel structural fracture acoustic emission signals $\boldsymbol{x} \in \mathbb{R}^{1\times T}$ manifest as single-channel temporal sequences. High sampling rates induce temporal sparsity whereas channel dimensionality lacks semantic expressiveness. Single-channel representation fails to capture multi-modal response characteristics of stress wave propagation. Lengthy temporal sequences introduce contradictions between computational complexity and real-time constraints. Consequently, initial feature extraction necessitates temporal-to-channel dimension transposition converting temporal resolution into channel semantic richness. Acting as the core component of the manifold projection mechanism, the Stem module serves as a learnable signal manifold projector mapping sparse single-channel sequences into compact multi-channel feature spaces. This mapping is achieved through three cascaded convolutional layers where each layer executes local temporal convolution and down-sampling operations. Kernel size and stride are strategically configured as 3 and 2 to capture local waveform structures of steel structural fracture transient pulses. For input signal $\boldsymbol{x} \in \mathbb{R}^{1\times T}$, the transformation process is formalized as three sequential composite functions.

$$\boldsymbol{F}_{\text{stem}} = \mathscr{H}_3 \circ \mathscr{H}_2 \circ \mathscr{H}_1(\boldsymbol{x}) \tag{1}$$

where each transformation unit $\mathscr{H}_i(\boldsymbol{z}) = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv1D}(\boldsymbol{z};\theta_i)))$ represents the i−th convolution-normalization-activation operation.

Three stride-2 downsampling operations compress temporal length to 1/8 of the original sequence whereas channel dimensionality progressively expands to initial feature dimension C_0. Output feature dimensionality becomes $\boldsymbol{F}_{\text{stem}} \in \mathbb{R}^{C_0 \times T/8}$. Progressive channel expansion and rapid sequence length reduction enable the Stem module to capture multi-scale features from high-frequency transient pulses to low-frequency envelope variations while reducing computational burden.

2.4. Hierarchical multi-stage architecture

Building upon initial features extracted by the Stem module, subsequent processing addresses the intrinsic hierarchical nature of steel structural fracture acoustic emission signals. Specifically, steel structural fracture acoustic emission signals exhibit multi-scale coupling characteristics where microscopic transient pulses carry high-frequency short-term waveform details whereas macroscopic evolution trends embody low-frequency long-term semantic patterns. Consequently, feature abstraction constructs a pyramidal architecture achieving hierarchical mapping from waveform details to semantic patterns through progressive downsampling and channel expansion. To this end, the backbone network adopts a three-stage cascaded structure where channel dimensionality and temporal length across stages exhibit inverse scaling with channels increasing whereas temporal length decreasing to maintain computational balance. Inter-stage embedding layers execute joint transformation of temporal compression and channel expansion. Formally, this transformation is expressed as

$$\boldsymbol{F}_{\text{emb}}^{(i)} = \text{ReLU}(\text{BN}(\text{Conv1D}(\boldsymbol{F}_{\text{out}}^{(i-1)}; k=3, s=2, p=1))) \tag{2}$$

where $\boldsymbol{F}_{\text{out}}^{(i-1)}$ denotes output features from the $(i-1)-$th stage. In this configuration, channel dimensionality C_i and temporal length L_i of the $i-$th stage satisfy $C_i = 2^i \cdot C_0$ and $L_i = L_0/2^{i-1}$ where C_0 denotes base channel number and L_0 denotes Stem output sequence length. Inverse scaling ensures computational complexity balance across stages. The pyramidal architecture aligns physical multi-scale characteristics with network hierarchy. Specifically, shallow stages capture microscopic waveform details on long sequences whereas deep stages extract macroscopic semantic patterns on compressed sequences. To further coordinate local transient feature preservation and global dependency modeling, each stage embeds an LDS-Former block. LDS-Former blocks decouple convolutional pathway and attention pathway through dual-stream architecture.

2.5. LDS-former block with dual-stream design

Building upon the pyramidal architecture achieving cross-stage multi-scale mapping, internal processing within each stage decouples local discriminability and global dependency modeling to synergistically capture dual characteristics of steel structural fracture acoustic emission signals. Convolutional operations excel at extracting fine waveform structures of local transient pulses, whereas attention mechanisms excel at modeling long-range dependencies of crack evolution but introduce quadratic computational complexity on long sequences. Consequently, LDS-Former blocks adopt dual-stream parallel architecture decoupling two modeling pathways. The local stream preserves original temporal resolution to capture complete waveform details of transient pulses. The global stream compresses temporal sequences to model long-range dependencies under computationally feasible constraints. Given input features $\boldsymbol{X}_i \in \mathbb{R}^{C_i \times L_i}$ of the $i-$th stage, LDS-Former blocks execute initial transformation through inverted residual blocks

$$\boldsymbol{X}_i^l = \mathscr{F}_{\text{InvRes}}^{m_i}(\boldsymbol{X}_i) \tag{3}$$

where $\mathscr{F}_{\text{InvRes}}^{m_i}(\cdot)$ denotes cascaded operations of inverted residual blocks. Inverted residual blocks leverage expand-then-compress channel transformation coupled with depthwise separable convolution to extract local waveform features while reducing parameter count to maintain lightweight design. Intermediate features $\boldsymbol{X}_i^l$ after initial transformation bifurcate into two parallel pathways. In the local stream, $\boldsymbol{X}_i^l$ transmits to fusion layers through pass-through connections preserving complete transient pulse information. Conversely, in the global stream, adaptive average pooling executes sequence compression, reducing temporal length to a fixed value L_g to circumvent quadratic complexity of attention mechanisms on lengthy sequences. This operation is formalized as

$$\boldsymbol{X}_i^{g,\text{pool}} = \text{AdaptiveAvgPool1D}(\boldsymbol{X}_i^l; L_g) \tag{4}$$

Compressed global features $\boldsymbol{X}_i^{g,\text{pool}} \in \mathbb{R}^{C_i \times L_g}$ undergo attention block processing in this latent space to capture long-range dependencies of crack evolution. This processing is expressed as

$$\boldsymbol{X}_i^{g,\text{attn}} = \mathscr{F}_{\text{Attn}}(\boldsymbol{X}_i^{g,\text{pool}}) \tag{5}$$

To achieve feature alignment with the local stream, the global stream upsamples to original temporal length L_i. This upsampling operation is expressed as

$$\boldsymbol{X}_i^g = \text{Upsample}(\boldsymbol{X}_i^{g,\text{attn}}; L_i) \tag{6}$$

Dual-stream fusion introduces gated modulation mechanism to adaptively determine weight allocation between transient pulses and evolution trends. Global stream features generate gated weight maps

$$\boldsymbol{W}_i^g = \text{Sigmoid}(\text{Conv1D}(\text{BN}(\boldsymbol{X}_i^g); k=1)) \tag{7}$$

Gated weight maps $\boldsymbol{W}_i^g$ modulate local stream features $\boldsymbol{X}_i^l$ to achieve adaptive feature selection. Modulated local features and global features are concatenated along channel dimension and fused through pointwise convolution

$$\boldsymbol{X}_i^u = [\boldsymbol{X}_i^l \odot \boldsymbol{W}_i^g, \boldsymbol{X}_i^g] \tag{8}$$

$$\boldsymbol{Y}_i = \text{Conv1D}(\text{BN}(\boldsymbol{X}_i^u); k=1, C_{\text{out}} = C_i) \tag{9}$$

where $\odot$ denotes Hadamard product and $[\cdot,\cdot]$ denotes channel-wise concatenation. Gated modulation mechanism endows the model with active selection capability to adaptively enhance contributions of transient pulses or evolution trends according to signal characteristics.

2.6. Attention mechanism with convolutional enhancement

The global stream models long-range dependencies in compressed temporal domain. To further enhance representational capacity, we propose convolution-enhanced attention mechanism to capture local temporal morphology of transient pulses within compressed sequences. Pure attention mechanisms lack inductive bias for local structure whereas transient pulses in acoustic emission signals inherently exhibit local waveform continuity. Consequently, we inject convolutional inductive bias into value vector mapping to preserve local temporal morphology characteristics. Additionally, acoustic emission wave propagation in steel structures manifests relative distance attenuation characteristics. We model this physical correlation through learnable relative positional bias. For each attention block input $\boldsymbol{X} \in \mathbb{R}^{C_i \times L_g}$ in the global stream, base representation is generated through shared pointwise convolution

$$\boldsymbol{Y}_{\text{base}} = \text{BN}(\text{Conv1D}(\boldsymbol{X}; k=1)) \tag{10}$$

The base representation $\boldsymbol{Y}_{\text{base}}$ is used to construct both queries and keys, while value vectors capture local temporal structure through an additional depthwise convolution path:

$$\boldsymbol{V}' = \text{DW - Conv}(\text{BN}(\boldsymbol{Y}_{\text{base}})) + \boldsymbol{Y}_{\text{base}} \tag{11}$$

where depthwise separable convolution preserves temporal morphology information of fracture pulses through local receptive fields. Subsequently, $\boldsymbol{Y}_{\text{base}}$ and $\boldsymbol{V}'$ are reshaped into multi-head forms $\boldsymbol{Q},\boldsymbol{K} \in \mathbb{R}^{B \times N_h \times L_g \times d}$ and $\boldsymbol{V} \in \mathbb{R}^{B \times N_h \times L_g \times d}$ for attention weight computation, where N_h denotes number of attention heads and $d = C_i/N_h$ represents dimension per head. Given that acoustic emission wave propagation paths in steel structures correlate closely with crack locations, attention computation incorporates learnable positional bias $\boldsymbol{b} \in \mathbb{R}^{1 \times N_h \times L_g \times L_g}$ to model such structured correlations:

$$\boldsymbol{A} = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^\top}{\sqrt{d}} + \boldsymbol{b}\right) \tag{12}$$

$$\boldsymbol{X}_{\text{attn}} = \text{Proj}(\boldsymbol{A}\boldsymbol{V}) + \boldsymbol{Y}_{\text{base}} \tag{13}$$

After projection layer restores attention output to C_i dimensions, a feed-forward network further provides nonlinear expressiveness:

$$\boldsymbol{X}_{\text{out}} = \text{FFN}(\boldsymbol{X}_{\text{attn}}) + \boldsymbol{X}_{\text{attn}} \tag{14}$$

where the feed-forward network $\text{FFN}(\boldsymbol{z}) = \text{Dropout}(\boldsymbol{W}_2 \cdot \text{ReLU}(\boldsymbol{W}_1 \boldsymbol{z}))$

employs a two-layer fully-connected structure with intermediate dimension $4C_i$, providing sufficient representational capacity while maintaining a lightweight design aligned with strict computational budgets.

2.7. Classification head

Through progressive feature abstraction via Stem module, hierarchical backbone network and LDS-Former blocks, LDS-Former captures hierarchical representations of steel structural fracture acoustic emission signals across multi-scale temporal-spatial domains. The local stream preserves fine waveform details of transient pulses, the global stream models long-range dependencies of crack evolution, whereas convolution-enhanced attention mechanism synchronously injects local inductive bias and physical correlation priors. These synergistically extracted representations output as deep feature maps rich in discriminative information at the final stage. The classification head serves as the terminal module of LDS-Former transforming hierarchical representations into category decisions for steel structural fracture states to support offshore platform service safety assessment. Given final stage output feature maps $\boldsymbol{Y}_3 \in \mathbb{R}^{C_3 \times L_3}$, the classification head first aggregates along temporal dimension into fixed-length vectors through global adaptive average pooling

$$\boldsymbol{z} = \text{AdaptiveAvgPool1D}(\boldsymbol{Y}_3; L_{\text{out}} = 1) \in \mathbb{R}^{C_3} \tag{15}$$

Feature vector $\boldsymbol{z}$ encodes global representation of the entire fracture process signal then maps to $K-$dimensional category logit space through fully-connected layer

$$\boldsymbol{o} = \boldsymbol{W}_{\text{fc}}\boldsymbol{z} + \boldsymbol{b}_{\text{fc}} \tag{16}$$

Category probability distribution is generated through Softmax function as $P(y = k|\boldsymbol{x}) = \exp(o_k)/\sum_{j=1}^{K}\exp(o_j)$. Category space $\mathscr{Y}$ encompasses pre-fracture states, fracture states across component types, and various crack configurations achieving fine-grained recognition of offshore steel structural platform service states. LDS-Former training employs cross-entropy loss function to optimize discrepancy between predicted probability distribution and ground-truth labels.

3. Experiment setup and dataset

3.1. Experimental configuration

A series of physical fatigue experiments was designed and conducted to construct an acoustic emission signal database encompassing diverse specimen geometries and damage patterns. Cyclic loading was applied via an Instron 3382 fatigue testing machine, while stress wave signals were acquired in real-time through a Physical Acoustics Corporation PCI-2 AE system throughout the damage evolution process. To establish a progressive validation framework, two representative specimen types were selected. First, 50 mm-thick DH36-Z35 marine-grade steel plates served as baseline specimens, providing uniform stress distribution and well-defined boundary conditions for reliable damage feature identification. Subsequently, Q345 low-alloy high-strength steel thin-walled beams with greater geometric complexity were introduced to simulate operational conditions of aging offshore platforms. The thin-walled beams adopted I-shaped cross-sections with dimensions of 100 mm height, 50 mm flange width, 2 mm web thickness, and 4 mm flange thickness.

For plate specimens, five structural states were systematically configured: intact state, central linear crack, central circular crack, eccentric linear crack, and eccentric circular crack. This configuration was designed to systematically examine the influence of crack geometry, spatial distribution, and stress field variations on acoustic emission characteristics. The thin-walled beams, owing to their thin-wall characteristics and open cross-sections, exhibit bending-shear coupling and local buckling under cyclic loading, producing wave propagation mechanisms fundamentally different from plates. For these specimens, three structural states were configured: intact state, eccentric linear crack, and eccentric circular crack. Across all specimen-crack combinations, acoustic emission signals were continuously recorded throughout the fatigue life, encompassing the complete progression from damage initiation through stable propagation to final fracture. This design ensures the dataset captures both pre-fracture signals characterizing cumulative damage evolution and fracture signals representing critical instability states.

Constant-amplitude cyclic tensile loading was applied via the Instron 3382 fatigue testing machine, with parameters differentiated according to load-bearing capacity and crack propagation rate requirements. Plate specimens were subjected to load cycles ranging from 16.67 kN to 166.67 kN with 75 kN amplitude at 10 Hz. Thin-walled beam specimens were subjected to load cycles from 6 kN to 60 kN with 27 kN amplitude at 5 Hz. Fig. 2 illustrates the experimental apparatus and specimen configurations.

The signal acquisition system utilized R15-type AE piezoelectric sensors with frequency response spanning 50 kHz to 400 kHz and resonance center frequency at 150 kHz. Signals were conditioned through 40 dB gain preamplifiers with amplitude recording range from 0 to 100 dB. To ensure temporal resolution and spectral accuracy for transient waveforms, the AEWin software was configured at 5 MHz sampling rate. Sensors were mounted onto specimen surfaces using acoustic couplant and magnetic fixtures to achieve low-distortion stress wave transmission.

3.2. Data preprocessing

Raw acoustic emission signals were categorized into 9 classes based on specimen type, crack configuration, and fracture status. Time-series samples of 1024 sampling points were extracted using non-overlapping sliding windows, and the dataset was partitioned via five-fold cross-validation. Table 1 presents class definitions and data partitioning details.

3.3. Evaluation metrics

Classification performance was assessed using accuracy and macro-averaged F1 score. For imbalanced multi-class tasks, macro-averaged F1 score computes the arithmetic mean of per-class F1 scores, providing an unbiased comprehensive performance measure. Statistical significance was validated through analysis of variance, computing performance means and standard deviations across five-fold cross-validation.

3.4. Computational setup

We conducted all computational experiments on a unified hardware-software platform with standardized training configurations to ensure fair comparison. The hardware platform features an Intel Core i9-14900HX processor, NVIDIA GeForce RTX 4070 Laptop GPU, and 32 GB memory. We constructed the software environment based on Python 3.9.19, PyTorch 2.0.0, CUDA 11.8, and cuDNN 8.7.0. Model training employed the Adam optimizer with initial learning rate of 0.001, batch size of 32, and maximum training epochs of 100. After training, we convert LDS-Former to Open Neural Network Exchange (ONNX) format and subsequently generate C++ inference code for microcontrollers through the STM32Cube.AI toolchain, with the complete LDS-Former deployment workflow illustrated in Fig. 3. The edge deployment target platform is the STM32H7B3I-DK development board, which features an ARM Cortex-M7 processor running at 280 MHz, 1.4 MB SRAM, and 2 MB flash memory, with detailed technical specifications presented in Table 2.

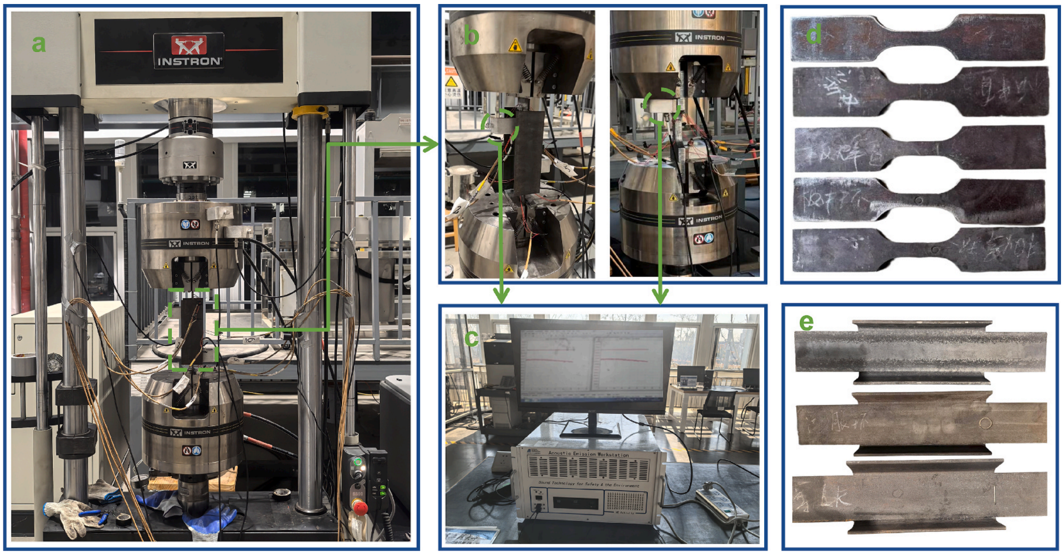


Fig. 2. Schematic diagram of the experimental setup and specimen configurations.

Table 1
Dataset labels and sample distribution.

Label	Specimen Type	Crack Configuration	Fracture Status	Training Set	Test Set
0	All specimens	All configurations	Non-fractured	5600	1400
1	Plate	No crack	Fractured	800	200
2	Plate	Central linear crack	Fractured	800	200
3	Plate	Central circular crack	Fractured	800	200
4	Plate	Eccentric linear crack	Fractured	800	200
5	Plate	Eccentric circular crack	Fractured	800	200
6	Thin-walled beam	No crack	Fractured	800	200
7	Thin-walled beam	Eccentric circular crack	Fractured	800	200
8	Thin-walled beam	Eccentric linear crack	Fractured	800	200

3.5. Implementation details

To facilitate understanding and reproducibility, this section outlines key implementation details of LDS-Former. Table 3 presents core parameters for model configuration and training settings. In terms of model configuration, we set the Stem output channels to 8 to ensure compact initial feature embeddings, with three-stage backbone channels configured as 16, 32, and 64 to create a hierarchical feature pyramid. Notably, in the value path of the global stream, we set the depthwise convolution kernel size to 3 to capture local continuity between adjacent sampling points, with both stride and dilation set to 1 to maintain temporal resolution, and padding set to 1 to ensure that the output length matches the input. Depthwise convolution is applied only to value vectors, not query or key vectors, based on the functional specialization within attention mechanisms: queries and keys compute similarity weights between positions through global comparison, independent of local temporal structure, while value vectors carry feature content whose local waveform morphology directly impacts output representation. Therefore, injecting convolutional inductive bias into the value path enhances local temporal structure sensitivity while preserving the ability to model global dependencies. Furthermore, considering the transient-evolutionary duality of AE signals, the Global Stream's adaptive pooling length is default to 16, which enforces the model to focus on low-frequency damage trends through aggressive compression. For training, we use the Adam optimizer to leverage its adaptive learning rate adjustment, with an initial learning rate of 0.001, a batch size of 32 to balance gradient stability and memory usage, a

Table 2
Technical specifications of STM32H7B3I-DK Discovery Kit.

Parameter	STM32H7B3I-DK Discovery Kit
MCU	STM32H7B3LIH6Q
CPU Core	ARM Cortex M7 (32-bit)
CPU Frequency	280 MHz
SRAM	1.4 MB
Flash	2 MB
FPU	✓
Supply voltage	3.3 V

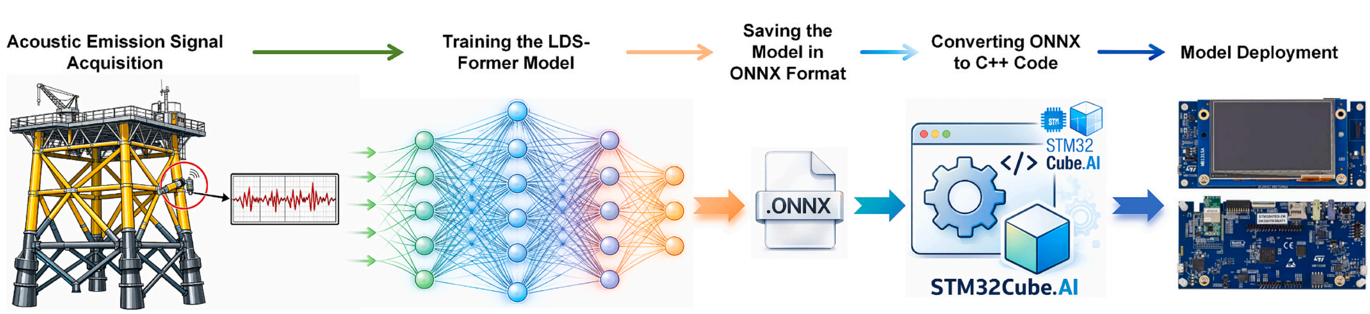


Fig. 3. LDS-Former Deployment Workflow.

Table 3
Implementation details of LDS-Former.

Category	Parameter	Value
Model Configuration	Stem Output Channels	8
	Stage Channels	[16, 32, 64]
Depthwise Convolution	Kernel Size	3
	Stride	1
	Padding	1
	Dilation	1
Training Configuration	Optimizer	Adam
	Learning Rate	0.001
	Batch Size	32
	Epochs	100
	Loss Function	Cross-Entropy

maximum of 100 training epochs for sufficient convergence, and cross-entropy loss to optimize the probability distribution for multi-class tasks.

3.6. Baseline methods

To comprehensively evaluate LDS-Former performance, this study establishes a two-tier baseline comparison framework. The first tier comprises classic architecture baselines including MobileNetV3, ResNet34, ShuffleNetV2, and Transformer. MobileNetV3 and ShuffleNetV2 represent different technical approaches to mobile-oriented lightweight design for benchmarking LDS-Former's lightweight characteristics. ResNet34 provides a strong baseline for local feature extraction as a classic deep convolutional network. Transformer adopts pure self-attention architecture with encoder composed of multi-head self-attention layers and feed-forward networks to benchmark LDS-Former's dual-stream architecture in global dependency modeling.

The second tier comprises domain-specific state-of-the-art baselines including MCSAT[43], TCACFormer[44], and LiConvFormer[45]. These three methods represent the latest lightweight Transformer variants rigorously validated in industrial intelligent operation and maintenance and published in top-tier journals. MCSAT achieves lightweight design through multi-scale convolutional sparse attention mechanism and pyramid convolution patch embedding module, dynamically focusing on key features via bi-level routing attention, demonstrating superior performance in rotating machinery fault diagnosis. TCACFormer combines multi-scale separable convolution with temporal-channel attention module, achieving efficient feature extraction through lightweight architecture, reaching leading-edge performance in gearbox and bearing fault identification. LiConvFormer combines separable multiscale convolution blocks with broadcast self-attention mechanisms, gaining widespread adoption in industrial equipment health monitoring. The introduction of these three methods validates LDS-Former's technical superiority relative to current state-of-the-art lightweight Transformer approaches in the intelligent O&M domain.

4. Experimental results and discussion

4.1. Signal characteristic analysis

This section characterizes time–frequency patterns of AE signals during steel structure damage evolution across different specimen types and crack configurations. Fig. 4 and Fig. 5 illustrate time-domain waveforms and spectral features respectively, with each row comparing a specific specimen-crack configuration between non-fractured and fractured states. Specifically, non-fractured signals exhibit low-amplitude diffuse noise and lack distinct spectral peaks. In sharp contrast, fractured signals display pronounced transient pulses accompanied by significant spectral energy concentration. Dominant frequency components are localized within the 0.15–0.25 normalized frequency interval, with peak amplitudes ranging from 60 to 200.

Plate central cracks produce sharp spectral main peaks, attributed to high energy concentration derived from symmetric stress fields. Conversely, eccentric cracks exhibit significantly lower energy concentration; asymmetric stress gradients excite mixed fracture modes, enhancing secondary peaks and broadening the frequency bandwidth. Furthermore, circular cracks exhibit richer secondary peak structures in high-frequency bands, stemming from interference effects of circumferential stress waves. Regarding thin-walled beams, they display distinctively different wave propagation characteristics. The I-shaped cross-section induces wave mode conversion, where multiple reflections extend signal duration and cause long-term reverberation. Consequently, spectral energy distribution becomes flatter relative to plates, and secondary peak amplitudes increase relatively, reflecting modulation effects of complex cross-sectional topology on wave propagation paths. Notably, spectral distinguishability among crack configurations diminishes in thin-walled beams compared to plates, indicating that geometric complexity attenuates the manifestation of frequency-domain crack characteristics.

4.2. Performance comparison with baseline models

To comprehensively validate the performance advantages of LDS-Former, this section systematically compares its performance with classic architectures and advanced lightweight Transformer intelligent diagnosis models through five-fold cross-validation. As illustrated in Fig. 6, comparative experiments quantitatively evaluate each model across diagnostic accuracy, training convergence, and computational efficiency. Results demonstrate that LDS-Former exhibits significant superiority in dual-tier benchmarks. Fig. 6(a) shows LDS-Former achieves 0.9505 accuracy, matching ShuffleNetV2 and significantly surpassing MobileNetV3′s 0.8750. Fig. 6(b) further demonstrates that LDS-Former surpasses advanced lightweight Transformer methods MCSAT's 0.8954, TCACFormer's 0.9110, and LiConvFormer's 0.8727, with relative improvements of 6.2%, 4.3%, and 8.9% respectively. ANOVA validates the statistical significance of these performance gains. Training loss curves in Fig. 6(c-d) show LDS-Former converges below 0.2 within 20 epochs with final loss stabilizing below 0.1, exhibiting superior convergence characteristics. These results confirm that LDS-Former achieves substantive breakthroughs relative to current state-of-the-art methods in the intelligent O&M domain.

Performance advantage stems from mechanism alignment between dual-stream architecture and physical characteristics of AE signals. AE signals couple high-frequency transient pulses and low-frequency evolutionary trends, demanding intelligent diagnosis models to simultaneously possess local discriminability and global dependency capturing capabilities. Classic lightweight networks fail to capture long-range dependencies due to local receptive field constraints. Transformer excels at global representation yet loses high-frequency information through tokenization. Advanced lightweight Transformer methods such as MCSAT and TCACFormer demonstrate superior performance in rotating machinery vibration signal processing, yet their designs target periodic stationary signals and exhibit limitations in synergistic modeling of transient mutations and long-range evolution in AE signals. In contrast, LDS-Former decouples functional specialization through dual-stream architecture, where the local stream preserves waveform details, the global stream captures long-range dependencies, and the gating mechanism adaptively fuses their contributions, achieving synergistic perception of transient-evolutionary dual characteristics.

Computational efficiency analysis reveals synergistic advantages of parameter reduction and accuracy improvement. As shown in Fig. 6(e), LDS-Former achieves ultra-lightweight design with 0.199 M parameters and 5.215 M FLOPs, accomplishing 97.2% parameter reduction and 98.5% computational cost reduction relative to ResNet34. Compared with MobileNetV3, LDS-Former improves accuracy by 8.6% while reducing parameters by 90%. This phenomenon reflects the optimal balance between parameter efficiency and feature expressiveness in

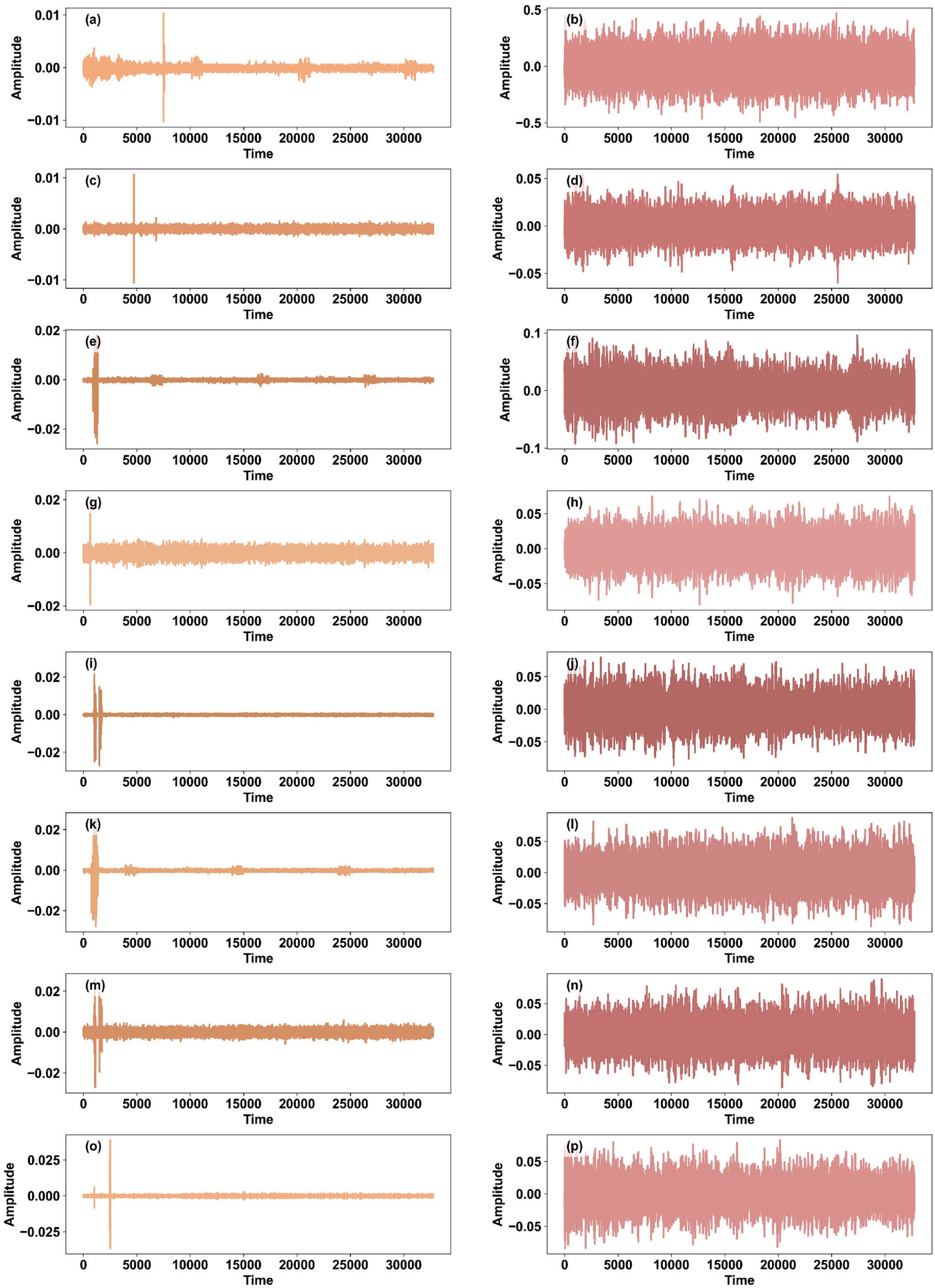


Fig. 4. Time-domain waveforms of acoustic emission signals. Each row compares non-fractured signals (left) versus fractured signals (right) for a specific specimen configuration: (a, b) Plate with no crack; (c, d) Plate with central linear crack; (e, f) Plate with central circular crack; (g, h) Plate with eccentric linear crack; (i, j) Plate with eccentric circular crack; (k, l) Thin-walled beam with no crack; (m, n) Thin-walled beam with eccentric circular crack; (o, p) Thin-walled beam with eccentric linear crack.

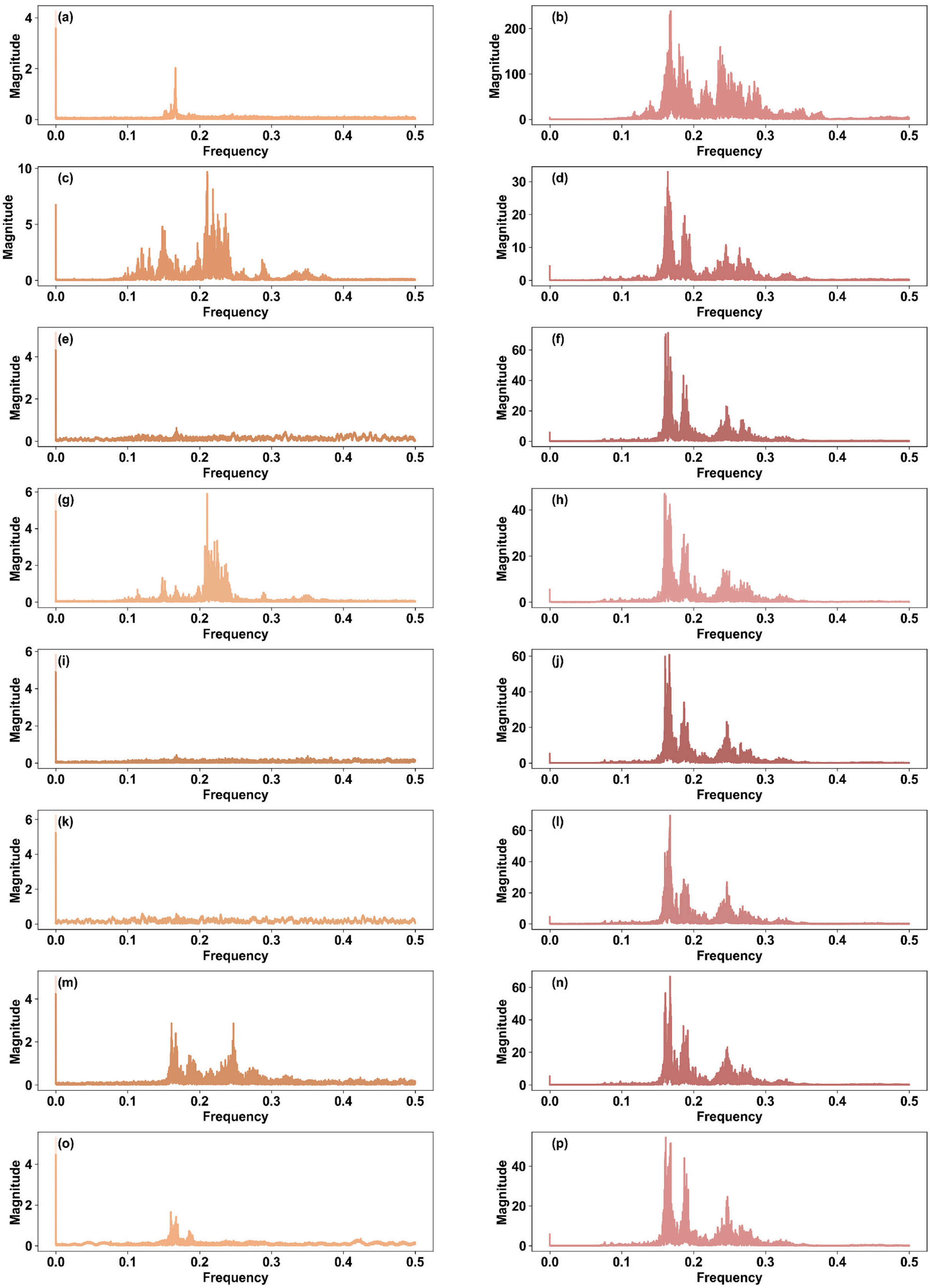
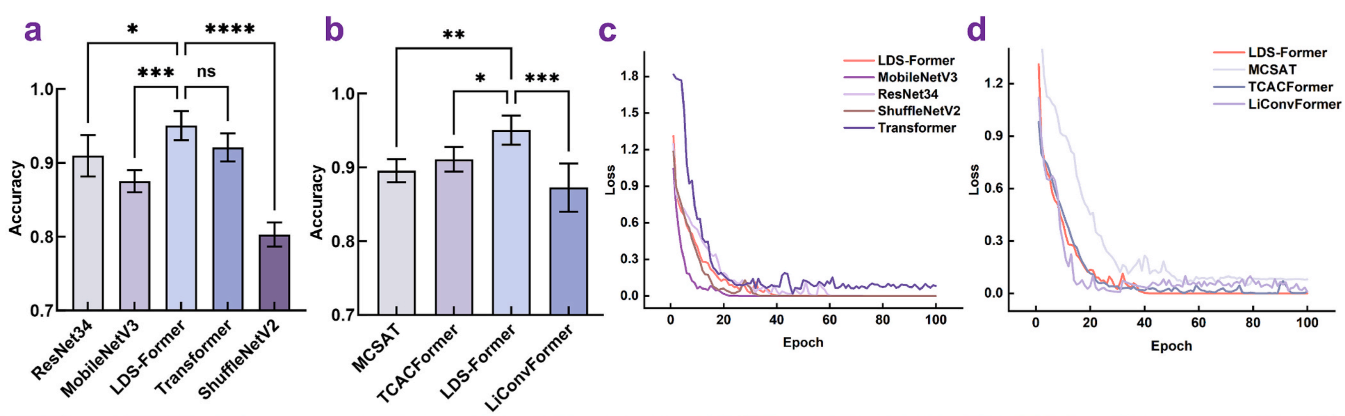


Fig. 5. Frequency-domain spectra of acoustic emission signals. Each row compares non-fractured signals (left) versus fractured signals (right) for a specific specimen configuration: (a, b) Plate with no crack; (c, d) Plate with central linear crack; (e, f) Plate with central circular crack; (g, h) Plate with eccentric linear crack; (i, j) Plate with eccentric circular crack; (k, l) Thin-walled beam with no crack; (m, n) Thin-walled beam with eccentric circular crack; (o, p) Thin-walled beam with eccentric linear crack.



e	LDS-Former	MobileNetV3	ResNet34	ShuffleNetV2	Transformer	MCSAT	TCACFormer	LiConvFormer
AVG. Accuracy	**0.9505**	0.8750	0.9096	0.9505	0.9209	0.8954	0.9110	0.8727
Params (M)	**0.199**	1.974	7.223	0.344	0.802	0.056	0.075	1.009
FLOPs (M)	**5.215**	20.490	357.356	13.074	806.626	10.197	10.735	44.813
Inference Time (ms)	**3.071**	4.471	3.243	4.245	2.407	1.826	2.293	5.873
Peak Memory Usage (MB)	**136.75**	143.91	39.53	137.65	18.08	12.67	136.66	140.25
Memory Bandwidth Utilization (GB/s)	**0.14**	1.39	8.01	0.19	1.26	0.17	0.15	1.19

Fig. 6. Performance comparison of LDS-Former with baseline models and state-of-the-art lightweight Transformer methods. (a) Accuracy comparison with classic architecture baselines. (b) Accuracy comparison with domain-specific state-of-the-art lightweight Transformer methods. (c) Training loss curves versus classic baselines. (d) Training loss curves versus state-of-the-art lightweight Transformer methods. (e) Comprehensive performance statistics table of all models, including average accuracy, parameter count, FLOPs, inference time, peak memory usage, and memory bandwidth utilization.

dual-stream architecture, where functional specialization of local and global streams enables the model to capture richer semantics with fewer parameters. Relative to MCSAT and TCACFormer, although LDS-Former increases parameters by 256% and 165% respectively, accuracy improves by 6.2% and 4.3%, indicating that additional parameters effectively translate into discriminative performance. Inference latency analysis further reveals nonlinear relationships between theoretical computational complexity and practical latency. Transformer's FLOPs reach 806.626 M, approximately 154 times that of LDS-Former, yet inference time is merely 2.407 ms, lower than LDS-Former's 3.071 ms. Similarly, MCSAT and TCACFormer's FLOPs are 10.197 M and 10.735 M respectively, approximately 2 times that of LDS-Former, yet inference times of 1.826 ms and 2.293 ms are both lower than LDS-Former. This phenomenon stems from memory access pattern differences. The attention mechanisms of Transformer, MCSAT, and TCACFormer execute regular matrix operations that facilitate GPU parallelization, with peak memory of 18.08 MB, 12.67 MB, and 136.66 MB respectively, and memory bandwidth utilization reaching 1.26 GB/s, 0.17 GB/s, and 0.15 GB/s respectively. In contrast, LDS-Former's dual-stream architecture involves heterogeneous operations such as feature bifurcation, compression, upsampling, and gated fusion, where different computational pathways constrain parallelization efficiency, resulting in peak memory of 136.75 MB and bandwidth utilization of only 0.14 GB/s. Despite latency disadvantage, LDS-Former's 3.071 ms inference time still meets real-time monitoring requirements, and its substantially reduced FLOPs and significantly improved diagnostic precision confer decisive advantages in resource-constrained edge device deployment.

4.3. Edge device deployment validation

To validate the deployment feasibility of LDS-Former on real edge devices, the trained LDS-Former is deployed onto the STM32H7B3I-DK development board. Table 4 presents measured deployment metrics. Inference time stabilizes at 2.086 ms with a fluctuation range of only 0.408 ms, meeting real-time requirements for online monitoring. Flash occupation of 845.0 KB and RAM occupation of 100.3 KB represent only 42.3% and 7.2% of the board capacity respectively, indicating that LDS-Former possesses ample deployment headroom on this platform. Since LDS-Former achieves native lightweight design through architectural optimization, weights of 767.97 KiB and activation memory of 48 KiB can be directly deployed at 32-bit floating-point precision, satisfying resource constraints without INT8 or FP16 quantization, thereby avoiding quantization-induced accuracy loss. In summary, edge deployment validation confirms that LDS-Former can run in real-time on resource-constrained microcontroller platforms, providing technical support for distributed intelligent monitoring on offshore platforms.

4.4. Feature representation analysis of dual-stream architecture

Building upon the overall performance validation, this section elucidates the intrinsic working mechanism of LDS-Former through visualization analysis of feature activation patterns. Fig. 7 depicts the activation patterns of the local stream, global stream, and fused features in the third stage, revealing the complementary nature of the dual-stream synergy. Specifically, the local stream features in Fig. 7(a) exhibit significant textural variations in the temporal dimension,

Table 4
Deployment Performance Metrics on STM32H7B3I-DK Edge Device.

Metric	Value
Mean Inference Time per Sample	2.086 ms
Inference Time Range	1.986–2.394 ms
Model Weight Size	767.97 KiB
Activation Memory	48 KiB
Total Flash Occupation	845.0 KB
Total RAM Occupation	100.3 KB

effectively capturing the high-frequency transient responses of fracture impulses. In sharp contrast, the global stream features in Fig. 7(b) demonstrate distinct temporal continuity along the time axis. This smooth morphology facilitates the modeling of low-frequency trends inherent in the long-range dependencies of crack evolution. Consequently, the fused features in Fig. 7(c) integrate these advantages through effective synthesis rather than simple superposition.

Channel-wise mean activation analysis further clarifies the feature selection process of the dual streams. As illustrated in Fig. 7(d), local stream activations display a high-variance distribution across channels, indicating strong feature selectivity. Conversely, Fig. 7(e) shows that the global stream maintains dense responses across channels but concentrates at significantly lower magnitudes compared to the local stream. This pattern suggests that the global stream functions primarily to provide a stable contextual background rather than dominating specific discriminative features. Finally, Fig. 7(f) demonstrates that the fused features inherit the high-variance distribution of the local stream while amplifying feature differences through inter-channel contrast enhancement, thereby improving the discriminative power for different damage states. Furthermore, Fig. 7(g) and (h) quantify the functional specialization of the two streams in the frequency domain: the local stream preserves much higher spectral energy in the mid- and high-frequency bands, whereas 96.2% of the global stream energy is concentrated in the low-frequency band with only small portions in the mid- and high-frequency bands. These results indicate that the local stream focuses on high-frequency details of transient pulses, while the global stream emphasizes low-frequency trends of damage evolution, providing frequency-domain evidence that the dual-stream architecture is consistent with the transient–evolutionary mechanism of AE signals.

To reveal the decision-making mechanism of the global stream, this section analyzes multi-head attention weight distributions. Fig. 8(a) through (d) present the weight matrices of the four attention heads in the third stage. Different heads display significant distribution variations across the compressed sequence. Head 1 disperses weights across the entire sequence with a peak of approximately 0.35. In contrast, Head 3 concentrates weights at specific positions with a peak reaching 0.7. Heads 2 and 4 fall between these patterns. These differences indicate that different heads perform distinct functions, enabling the global stream to simultaneously attend to overall trends and key features. Furthermore, we computed attention entropy to quantify the degree of selectivity. As shown in Fig. 9, Head 3 achieves a minimum entropy of 1.18 nats, while Head 2 maintains high entropy of 2.4 to 2.6 nats. Low entropy indicates attention concentrated on a few positions, whereas high entropy corresponds to uniform distribution. The entropy curves of different heads exhibit complementary patterns along the position dimension, jointly covering the entire compressed sequence.

Fig. 8(e) through (h) present the learnable positional bias matrices. Bias values range from -3 to $+3$, modulating attention weights through addition before Softmax. Unlike the dispersed or concentrated patterns of weights, bias exhibits clear structural features in space rather than random distribution. Positive bias enhances attention at corresponding position pairs, while negative bias weakens attention. Observations reveal that bias displays continuous positive or negative blocks in local regions rather than scattered points. This spatial continuity indicates that the model learns systematic correlation patterns between position pairs through training. Different heads show distinct bias topologies that correspond to their weight distribution characteristics, jointly determining the final attention allocation strategy.

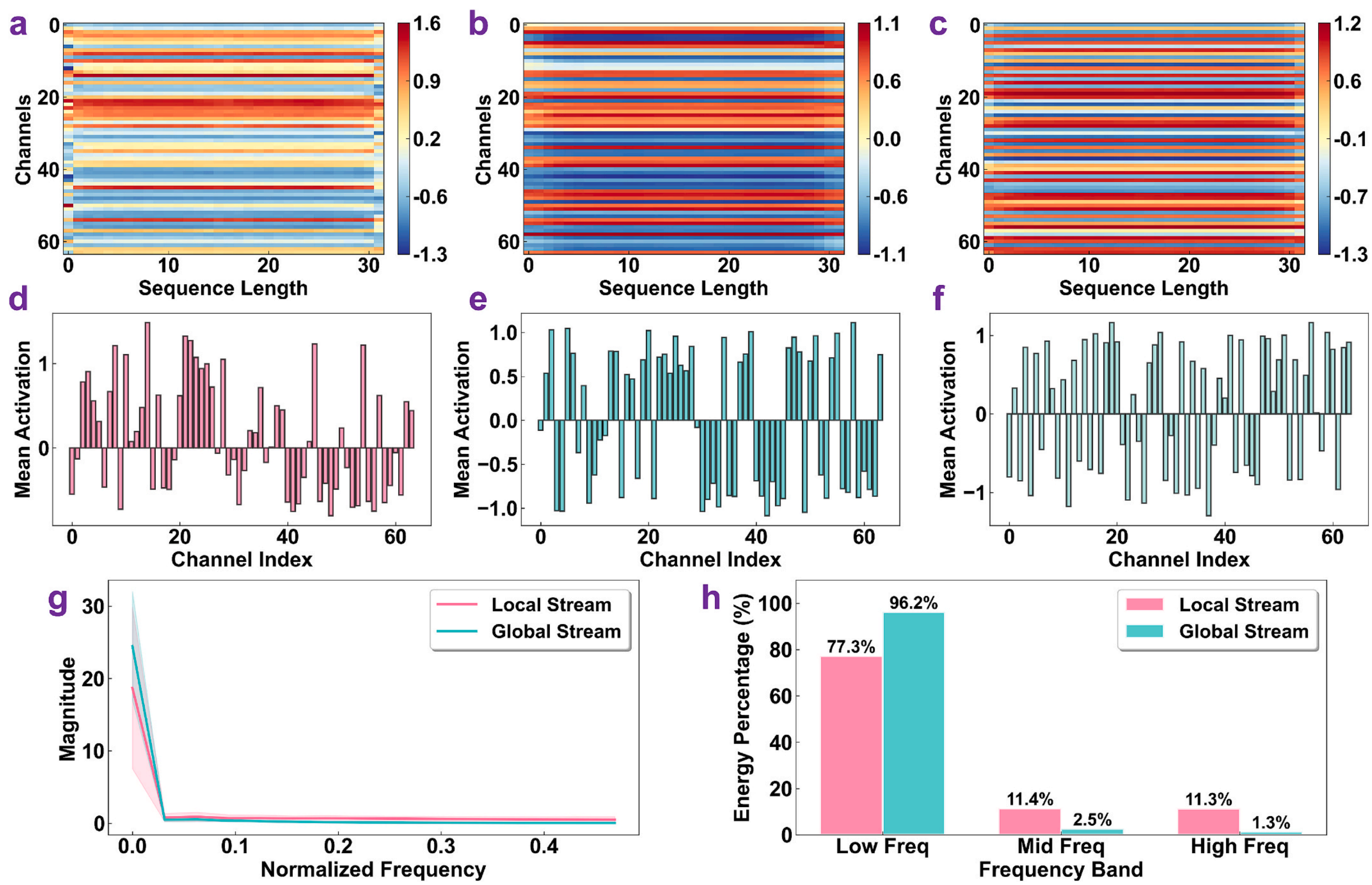


Fig. 7. Feature activation visualization of dual-stream architecture in the third stage: (a) Local stream features, (b) Global stream features, (c) Fused features, (d) Mean activation per channel for Local stream, (e) Mean activation per channel for Global stream, (f) Mean activation per channel for fused features, (g) Magnitude spectrum comparison of local and global streams, (h) Frequency band energy distribution of local and global streams.

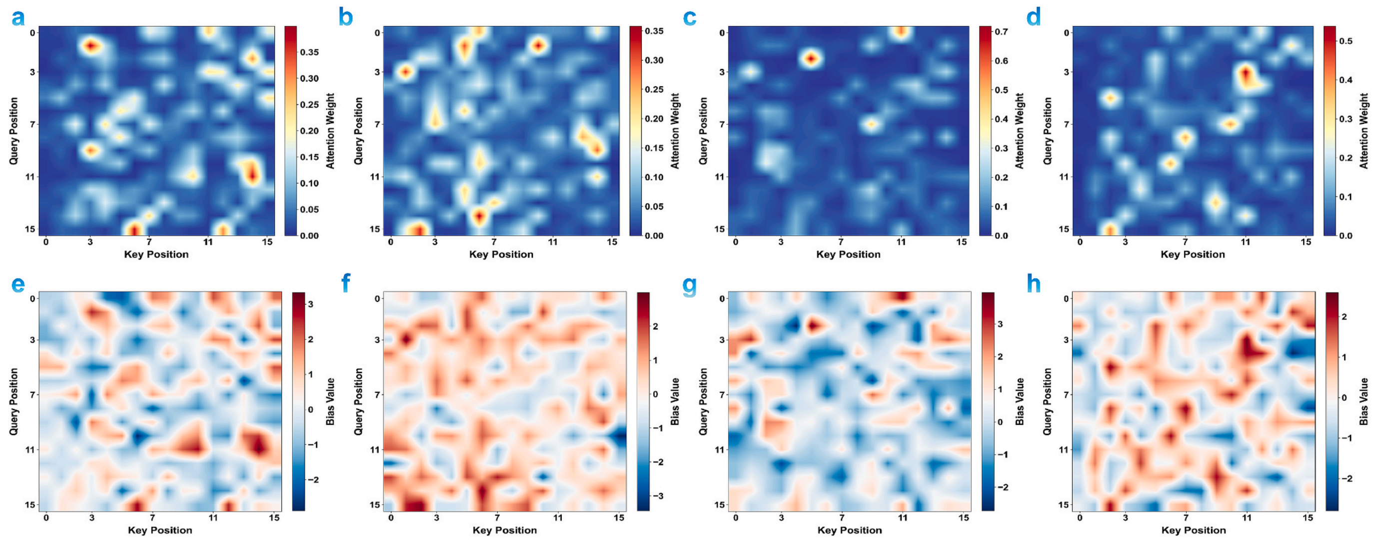


Fig. 8. Multi-head attention weight distributions and learnable positional bias in the third stage: (a) Attention weights of Head 1, (b) Attention weights of Head 2, (c) Attention weights of Head 3, (d) Attention weights of Head 4, (e) Positional bias of Head 1, (f) Positional bias of Head 2, (g) Positional bias of Head 3, (h) Positional bias of Head 4.

Fig. 9. Attention entropy across query positions for the four attention heads.

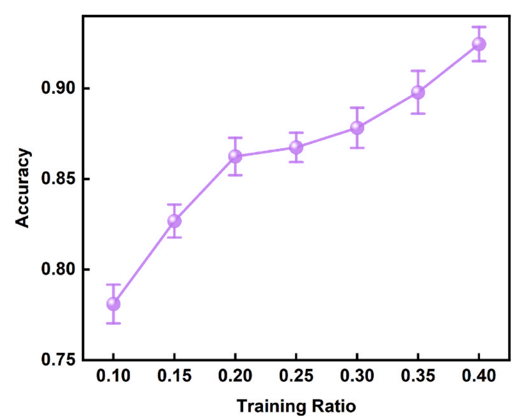


Fig. 10. LDS-Former Performance with Limited Training Data.

4.5. Performance under limited training data

Following the analysis of LDS-Former's feature representation mechanisms, this section further evaluates the generalization capability of LDS-Former under data-scarce scenarios, which constitutes a critical constraint in practical engineering deployment. Fig. 10 depicts accuracy variation of LDS-Former across different training ratios. Under severely constrained conditions with a minimal training ratio of 0.1, the model achieves an accuracy of 0.7810, demonstrating that LDS-Former effectively captures discriminative AE signatures even in extremely small-sample scenarios. As the training ratio increases to 0.2, accuracy surges to 0.8624 with an improvement of 10.42%, underscoring the efficacy of the dual-stream architecture in exploiting incremental training data. With the ratio extended to 0.4, accuracy further improves to 0.9244, corroborating the model's data-efficient learning capability.

To quantify the statistical reliability of performance trends, Table 5 presents statistical metrics derived from 5 independent experiments. Standard deviations remain below 0.012 across all training ratios, with coefficients of variation maintained within 1.4%, confirming stability and reproducibility of model outputs. At TR = 0.2, the 95% confidence interval of [0.8496, 0.8752] exhibits a narrow width of 0.0256, providing robust statistical support for the 86.2% accuracy. As TR increases to 0.4, the confidence interval converges to [0.9127, 0.9361], validating the model's performance advantage under relatively sufficient data. However, statistical analysis simultaneously reveals inherent limitations of supervised learning. At TR below 0.1, accuracy drops to 78.1%, showing significant degradation relative to TR = 0.2, which underscores the current architecture's dependence on labeled samples. Addressing this limitation, future research could incorporate semi-supervised or self-supervised learning techniques to enhance feature representation robustness by exploiting intrinsic structures in unlabeled AE data, thereby maintaining diagnostic capability under label-scarce conditions. In summary, LDS-Former achieves satisfactory diagnostic precision without requiring large-scale labeled data within the current supervised framework, providing a viable solution for labor-intensive offshore platform monitoring.

4.6. Robustness to noise contamination

To validate the robustness of LDS-Former under harsh operating conditions in marine environments, this section evaluates model diagnostic performance under strong noise interference. Offshore platform monitoring faces complex noise sources such as electromagnetic interference and sensor signal attenuation, rendering noise resistance a critical metric for practical deployment. We construct two representative noise scenarios for testing. First, salt-and-pepper noise simulates

Table 5
Statistical Analysis of Model Performance Under Different Training Ratio.

	TR = 0.1	TR = 0.15	TR = 0.2	TR = 0.25	TR = 0.3	TR = 0.35	TR = 0.4
Mean	0.7810	0.8268	0.8624	0.8674	0.8782	0.8978	0.9244
Std	0.0107	0.0091	0.0103	0.0081	0.0111	0.0118	0.0094
95%CILB	0.7677	0.8154	0.8496	0.8573	0.8644	0.8831	0.9127
95%CIUB	0.7944	0.8381	0.8752	0.8775	0.8920	0.9125	0.9361
CV	1.3750%	1.1040%	1.1960%	0.9338%	1.2680%	1.3150%	1.0190%

Note: “Std” refers to the Standard Deviation, “95% CILB” denotes the 95% Confidence Interval Lower Bound, “95% CIUB” stands for the 95% Confidence Interval Upper Bound, and “CV” represents the Coefficient of Variation.

random amplitude mutations caused by electromagnetic pulse interference and transient sensor failures, controlled by noise intensity parameter NI representing the proportion of contaminated sampling points. Second, pink noise simulates 1/f spectral noise introduced by low-frequency background vibrations in marine environments and mechanical equipment operation, controlled by signal-to-noise ratio SNR governing noise interference strength. As shown in Fig. 11, LDS-Former demonstrates excellent anti-interference capability under both noise types. Specifically, Fig. 11(a) shows that under salt-and-pepper noise interference, as noise intensity NI increases from 0.01 to 0.1, accuracy decreases from 0.9441 to 0.8485, with a relative decline of only 10.1%. Even at extreme noise intensity NI = 0.1, the model maintains 84.85% accuracy, validating the robustness of synergistic feature extraction between local and global streams. This phenomenon stems from the complementary characteristics of the dual-stream architecture. Although high-frequency transient pulses captured by the local stream are affected by salt-and-pepper noise, the global stream focuses on low-frequency evolutionary trends through sequence compression and attention mechanisms, remaining insensitive to random amplitude mutations, thus enabling gated fusion to suppress noise contamination effects. Furthermore, Fig. 11(b) presents performance variation under pink noise interference. As SNR decreases from 30 dB to 5 dB, accuracy decreases from 0.9548 to 0.8808, with a relative decline of only 7.8%. In strong noise scenarios at SNR = 5 dB, the model still maintains 88.08% accuracy. This noise resistance benefits from the inverted residual convolution structure of the local stream preserving discriminative waveform details, while the compressed-sequence attention of the global stream enhances tolerance to noise perturbations through long-range dependency modeling. Synthesizing results from both noise tests, LDS-Former maintains stable diagnostic performance under strong noise environments, validating its practical deployment feasibility under harsh operating conditions on marine platforms.

4.7. Hyperparameter sensitivity analysis

To establish the rationality of LDS-Former architectural configurations and verify its robustness against hyperparameter perturbations, this section systematically evaluates the impact of block quantity and learning rate on performance and efficiency. Block quantity ablation reveals diminishing marginal returns in performance gains. As illustrated in Fig. 12(a), accuracy surges from 0.83 to 0.95 as block quantity increases from 1 to 3, achieving a 14.5% gain. However, beyond 3 blocks, the performance curve saturates, yielding merely 0.01 marginal gain from 3 to 6 blocks, while Fig. 12(c) shows parameter count and inference latency escalate by 3.3 × and 6.6 × respectively. This imbalanced cost-benefit ratio demonstrates that the three-block configuration achieves optimal trade-off between accuracy and computational cost, ensuring optimal discriminability under lightweight constraints. Learning rate sensitivity analysis validates the optimization stability of the model. As depicted in Fig. 12(b), when learning rate fluctuates within the 0.001 to 0.01 interval, accuracy consistently maintains above 0.95 with standard deviation below 0.003, manifesting pronounced insensitivity to hyperparameter perturbations. Furthermore, Fig. 12(d) demonstrates that LDS-Former exhibits strong robustness to input signal length, maintaining accuracy above 0.93 with degradation below 0.02 as signal length extends from 1024 to 5120. The robustness to both learning rate and signal length collectively confirms that LDS-Former exhibits low dependence on training configurations and flexibly adapts to varying signal lengths in field conditions, reducing tuning costs and enhancing system reliability in practical deployment.

Regarding the selection of sequence compression length in the global stream, we systematically evaluate the impact of different pooling lengths on performance and efficiency. As shown in Fig. 13 and Table 6, when pooling length L_g increases from 16 to 128, accuracy decreases from 0.9598 to 0.8858 while parameter count increases from 0.199 M to 0.329 M. Spectral analysis reveals the physical mechanism of performance degradation. Under the $L_g = 16$ configuration, the compressed spectrum closely matches the pre-compression spectrum in the low-

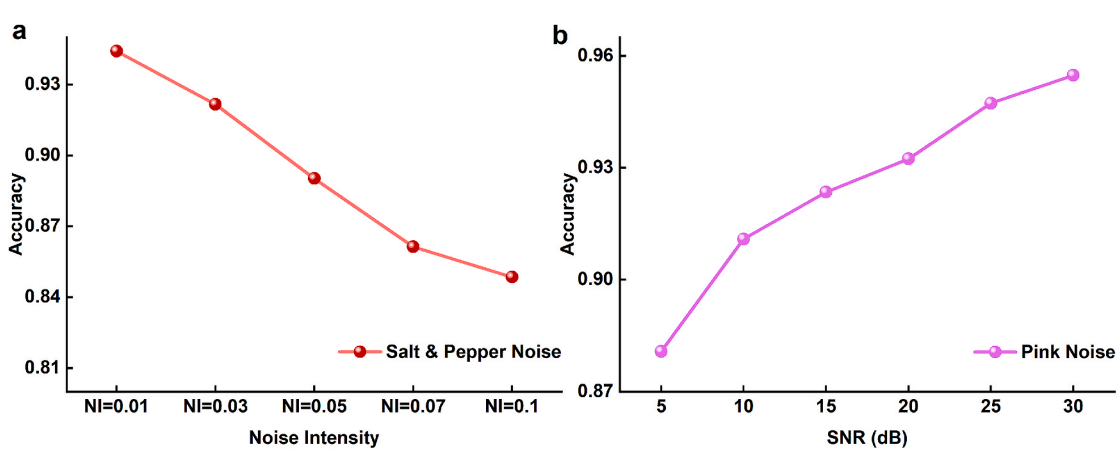


Fig. 11. Noise robustness evaluation of LDS-Former. (a) Performance degradation under salt-and-pepper noise with varying noise intensity (NI). (b) Performance variation under pink noise with different signal-to-noise ratios (SNR).

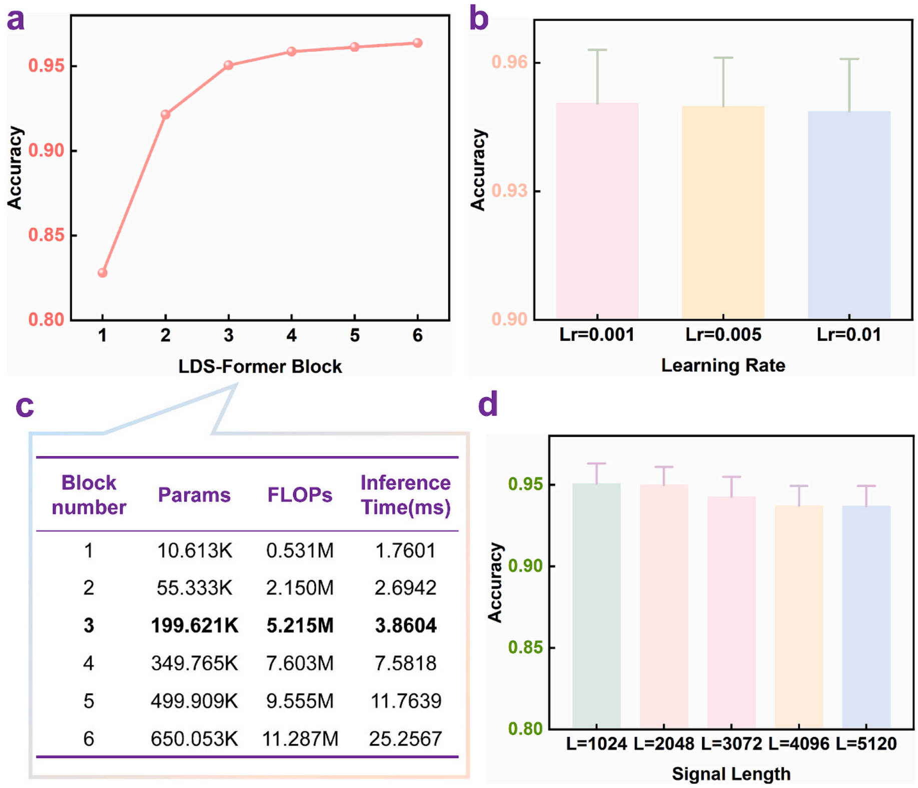


Fig. 12. Hyperparameter sensitivity analysis. (a) Accuracy versus number of LDS-Former blocks. (b) Accuracy across different learning rates. (c) Computational complexity and inference time versus number of LDS-Former blocks. (d) Accuracy across different input signal lengths.

frequency band below normalized frequency 0.1, while rapidly attenuating in high-frequency bands with amplitude significantly lower than pre-compression levels, which reflects the global stream's capability to effectively filter out high-frequency transient mutations while preserving macroscopic evolutionary trends. The corresponding spectrograms further confirm this mechanism, where the pre-compression spectrogram exhibits dense high-frequency energy fluctuations along the time axis, whereas the post-compression spectrogram smooths these transient details and retains only major energy concentration regions along the frequency axis, indicating that the global stream achieves focus on macroscopic patterns of damage evolution by ignoring local high-frequency mutations. Conversely, when L_g increases to 128, the compressed spectrum shows significantly reduced amplitude across the entire frequency range, with degraded time–frequency resolution in spectrograms and diffuse energy distribution, indicating that insufficient compression introduces redundant information and weakens the attention mechanism's ability to focus on critical evolutionary patterns. Integrating performance data with frequency-domain analysis, $L_g = 16$ maximizes computational efficiency while preserving discriminative low-frequency semantics, achieving optimal balance between performance and efficiency.

4.8. Ablation study

This section systematically dissects the independent contribution of each core component to the overall performance of LDS-Former through ablation experiments. To decouple the roles of the local stream, global stream, and gated fusion mechanism, four experimental configurations were designed. The baseline configuration is the full LDS-Former model, retaining all design components. First, by excluding the local stream and its inverted residual convolution pathway while preserving only the global stream's compressed-sequence attention mechanism, the necessity of local transient feature extraction is evaluated. Second, by omitting the global stream and its attention computation while maintaining only the local stream's convolutional pathway, the contribution of long-range dependency modeling is quantified. Third, dual streams are retained while discarding the gated modulation mechanism, employing simple channel-wise concatenation followed by pointwise convolution fusion, to verify the effectiveness of adaptive feature selection. To ensure a fair comparison, all ablation experiments were conducted on identical dataset partitions, training configurations, and hardware platforms, computing average accuracy and standard deviation through five-fold cross-validation.

As shown in Table 7, ablation experimental results elucidate the indispensability of each core component in LDS-Former. The absence of the local stream induces the most pronounced performance deterioration, with accuracy substantially declining from 0.9505 to 0.8893, representing a relative drop of 6.44%, confirming the critical role of the inverted residual convolution pathway in encoding fine-grained waveform details of steel structural fracture transient pulses. In contrast, the omission of the global stream reduces accuracy to 0.9127, representing a relative decline of 3.98%, underscoring the necessity of the compressed-sequence attention mechanism for modeling long-range dependencies in crack evolution, albeit with lesser impact than the local stream. The exclusion of the gated fusion mechanism results in accuracy declining to 0.9318, representing a relative drop of 1.97%, and this moderate degradation demonstrates that the adaptive feature selection strategy effectively coordinates weight allocation between local discriminability and global context, thereby optimizing dual-stream synergy.

To further validate the design rationale of applying depthwise convolution exclusively to the value path in the global stream, we design targeted ablation experiments to systematically evaluate the impact of different convolution configurations on attention mechanisms. Experiments compare three configuration schemes where the first is the baseline configuration adopting the proposed method that applies depthwise convolution only to the value path while query and key paths use pure linear projection. The second is the no-convolution

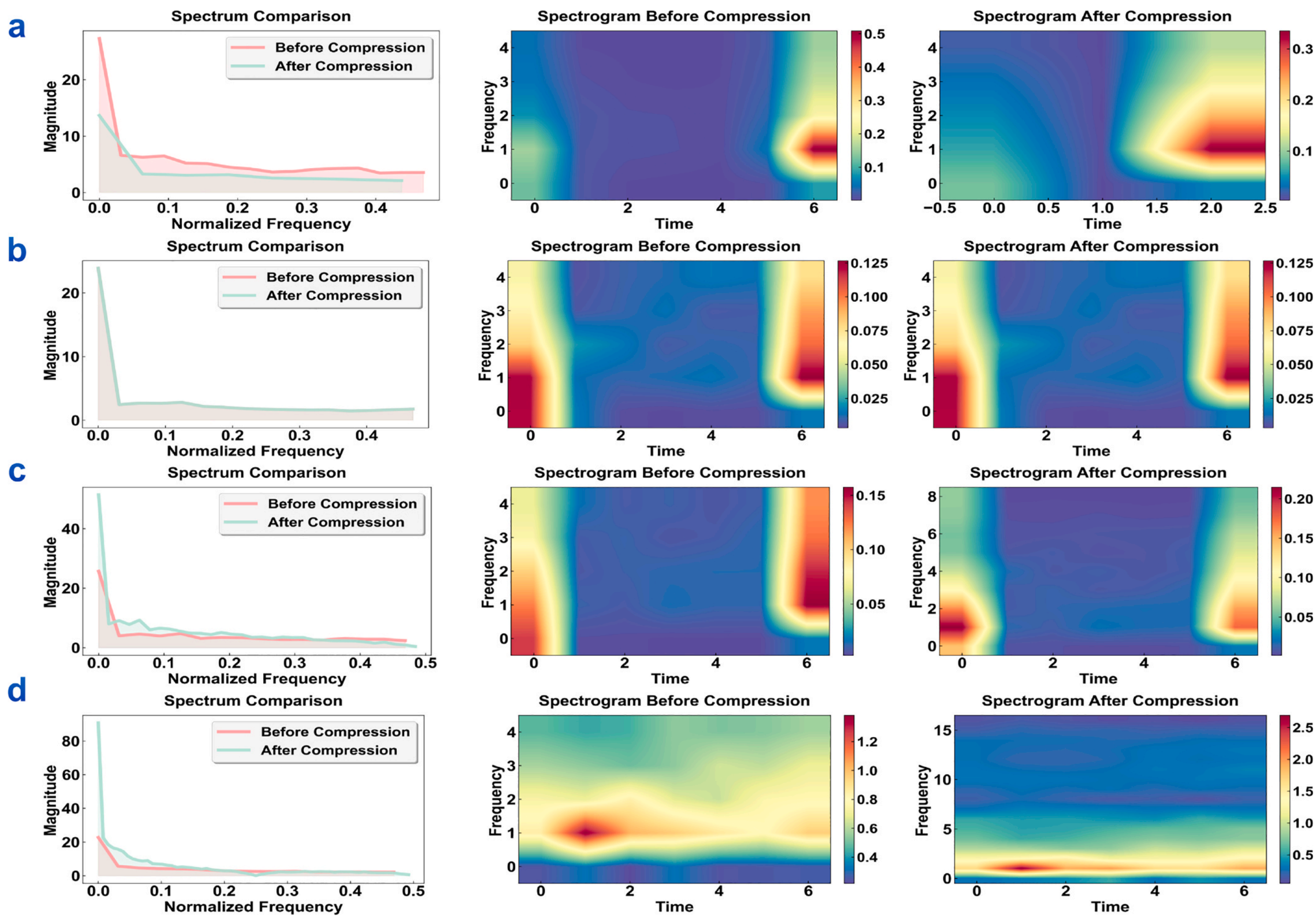


Fig. 13. Spectrum and spectrogram comparison before and after compression at different pooling lengths (a) L_g=16, (b) L_g=32, (c) L_g=64, (d) L_g=128.

Table 6
Performance and efficiency comparison across different pooling lengths.

Pooling Length	Accuracy	F1-Score	Parameters (M)
16	0.9598	0.9608	0.199
32	0.9360	0.9363	0.206
64	0.9184	0.9202	0.230
128	0.8858	0.8827	0.329

Table 7
Ablation study results on dual-stream architecture components.

Configuration	Local Stream	Global Stream	Gating Mechanism	Accuracy (Mean ± Std)	Relative Change
Full Model	✓	✓	✓	0.9505 ± 0.0082	/
w/o Local Stream	×	✓	×	0.8893 ± 0.0124	−6.44%
w/o Global Stream	✓	×	×	0.9127 ± 0.0095	−3.98%
w/o Gating Mechanism	✓	✓	×	0.9318 ± 0.0089	−1.97%

configuration where query, key, and value paths all adopt pure linear projection to verify the necessity of depthwise convolution for local temporal structure modeling. The third is the all-convolution configuration where query, key, and value paths all apply depthwise convolution to verify the rationality of adding convolution only to the value path. As shown in Table 8, the baseline configuration achieves 95.05% accuracy, the no-convolution configuration accuracy drops to 90.46% with a relative decline of 4.83%, indicating that depthwise convolution plays an important role in local temporal structure modeling. However, the all-convolution configuration accuracy reaches only 94.73%, declining 0.34% compared to the baseline configuration, revealing the negative effects of excessive introduction of convolutional inductive bias. This phenomenon stems from convolution operations in query and key paths interfering with the global comparison characteristics of position-wise similarity computation, weakening the attention mechanism's capability to capture long-range dependencies, whereas convolution enhancement in the value path preserves local morphology of feature content without affecting similarity measurement, thereby achieving optimal balance between local structure preservation and global dependency modeling. Experimental results confirm that the design choice of applying depthwise convolution exclusively to the value path possesses sufficient empirical support and theoretical justification.

Table 8
Ablation study on depthwise convolution placement in attention mechanism.

Configuration	Q Path	K Path	V Path	Accuracy	Relative Change
Baseline (Proposed)	Linear	Linear	DW-Conv	0.9505	0%
No Convolution	Linear	Linear	Linear	0.9046	−4.83%
All Convolution	DW-Conv	DW-Conv	DW-Conv	0.9473	−0.34%

4.9. Frequency-domain validation of mechanism-aligned design

This section validates the alignment between frequency-domain functional specialization of the dual-stream architecture and the physical duality of acoustic emission signals through spectral analysis. We extract intermediate-layer features from local and global streams in the trained model weights and compute power spectral density, spectral centroid, and high-frequency energy distribution. Fig. 14 presents four key frequency-domain measurements. Fig. 14(a) shows logarithmic-scale power spectra where the local stream maintains high energy across the entire frequency range while the global stream rapidly attenuates in high-frequency bands. Fig. 14(b) presents spectral centroid comparison where the bar chart clearly demonstrates that the local stream spectral centroid is substantially higher than the global stream. Fig. 14(c) displays normalized power spectra where the local stream maintains stable response across broad frequency range while the global stream concentrates in low-frequency bands. Fig. 14(d) shows high-frequency energy ratio where the local stream preserves significant high-frequency energy while the global stream high-frequency energy approaches zero. These measurements indicate that the local stream effectively preserves high-frequency components of transient pulses while the global stream focuses on capturing low-frequency evolutionary trends, confirming that the dual-stream design achieves frequency-domain alignment with the transient-evolutionary physical mechanism of acoustic emission signals.

4.10. Discussion on comparison with traditional signal processing methods

Compared with traditional acoustic emission signal processing methods, the core advantage of LDS-Former lies in the end-to-end learning paradigm that eliminates dependence on manual feature engineering. Traditional methods adopt multi-stage pipelines that extract time–frequency domain features through wavelet transform or empirical mode decomposition, subsequently select feature subsets based on expert knowledge, and finally feed features into shallow classifiers such as support vector machines or random forests. However, this pipeline exhibits three limitations. First, feature extraction relies on prior knowledge of signal processing experts and thus generalization capability is constrained to specific damage patterns. Second, handcrafted features are often optimized for single physical assumptions and struggle to adapt to the transient-evolutionary duality of AE signals, potentially missing critical discriminative patterns. Third, error accumulation introduced by multi-stage processing leads to suboptimal overall performance. In contrast, LDS-Former directly learns hierarchical feature representations from raw time-domain waveforms and outputs damage categories without manual feature design. Specifically, the dual-stream architecture automatically discovers discriminative patterns of local transient pulses and global evolutionary trends through data-driven approaches, while the gating fusion mechanism adaptively balances their contributions, thereby avoiding feature selection bias in traditional methods. Furthermore, end-to-end training optimizes the entire processing chain through backpropagation, enabling feature extraction and classification decisions to learn synergistically and significantly improving diagnostic accuracy. Experimental results validate this advantage, where LDS-Former achieves 86.24% accuracy with only 20% training data, demonstrating data efficiency and cross-component generalization capability that traditional methods struggle to match.

5. Conclusion

The LDS-Former method has been proposed to resolve the intrinsic conflict between lightweight constraints and accurate feature extraction in steel structural crack propagation recognition. This method targets the transient-evolutionary duality of acoustic emission signals by adopting a mechanism-aligned dual-stream architecture to achieve synergistic capture of heterogeneous features, where the local stream preserves waveform details of high-frequency transient pulses through inverted residual convolutions while the global stream models temporal dependencies of low-frequency evolutionary trends through compressed-sequence attention, and the gated fusion mechanism adaptively weights their contributions to optimize feature representation. To validate the effectiveness of this method, physical fatigue experiments have been designed and a dataset encompassing diverse

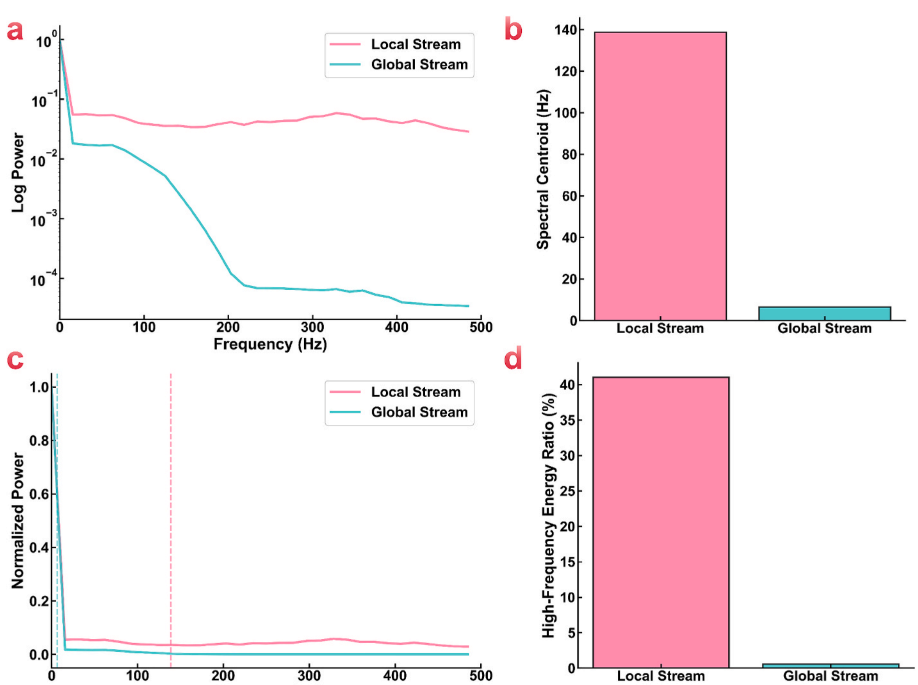


Fig. 14. Frequency-domain validation of dual-stream mechanism alignment. (a) Logarithmic-scale power spectra. (b) Spectral centroid comparison. (c) Normalized power spectra. (d) High-frequency energy ratio.

specimen geometries and complete damage lifecycle has been constructed. LDS-Former attains 95.05% accuracy while maintaining merely 0.199 M parameters and 5.215 M FLOPs, achieving 97.2% parameter reduction relative to deep architectures without sacrificing performance. Edge deployment validation demonstrates that LDS-Former achieves 2.086 ms inference latency on the STM32H7B3I-DK microcontroller platform with flash occupation of only 845.0 KB and RAM occupation of only 100.3 KB, confirming its deployment feasibility on resource-constrained embedded devices. Robustness evaluation further reveals industrial application potential where LDS-Former achieves 86.24% accuracy with only 20% training data, maintains stable performance across learning rates from 0.001 to 0.01, and adapts to signal length variations under different sampling configurations. Noise resistance testing shows that under extreme conditions of salt-and-pepper noise intensity 0.1 and pink noise signal-to-noise ratio 5 dB, LDS-Former still maintains 84.85% and 88.08% accuracy respectively, validating the tolerance of the dual-stream synergistic extraction mechanism to environmental interference. These characteristics indicate that LDS-Former exhibits significant advantages in data efficiency and hyperparameter robustness, contributing to reduced deployment and maintenance costs for offshore platform monitoring. Future research can introduce semi-supervised or self-supervised learning techniques to enhance feature representation capability by exploiting intrinsic structures in unlabeled acoustic emission data, thereby further improving diagnostic performance under label-scarce conditions and providing new technical pathways for real-time edge-side intelligent monitoring of steel structural damage under resource-constrained environments.

CRediT authorship contribution statement

Yuchen Lu: Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Zhenhao Zhu:** Validation. **Yifei Li:** Writing – review & editing. **Hongbing Liu:** Supervision, Funding acquisition. **Jiufan Hou:** Data curation. **Chunyu Zhou:** Data curation. **M.Abdel Wahab:** Writing – review & editing, Supervision.

Funding

This research is funded by the Joint Funds of the National Natural Science Foundation of China (U25A20372), the Key R&D Program of Shan- dong Province, China (2025CXGC010704), the Natural Science Foundation of Shandong Province of China (ZR2022QE091).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] J. Cui, C. Lv, J. Du, Real-time structural health monitoring of steel structures using acoustic emission signals and a KAN-LSTM deep learning framework, Eng. Struct. 344 (2025) 121328.
[2] J. Cui, C. Lv, X. Qu, J. Du, H. Wang, Development of an intelligent CNN-LSTM-attention model for acoustic emission-based fracture detection and structural health monitoring in marine steel structures, Ocean Eng. 339 (2025) 122002.
[3] Y. Sumi, Fatigue crack propagation in marine structures under seaway loading, Int. J. Fatigue 58 (2014) 218–224.
[4] Y. Zhang, X. Huang, F. Wang, Fatigue crack propagation prediction for marine structures based on a spectral method, Ocean Eng. 163 (2018) 706–717.
[5] Y. Zhong, Y. Shao, X. Gao, X. Luo, H. Zhu, Fatigue crack growth of EH36 steel in air and corrosive marine environments, J. Constr. Steel Res. 210 (2023) 108104.
[6] Y. Li, T. Hu, Q. Li, Y. Wu, L. Wang, Y. You, B. Wang, Evaluation of the stress corrosion crack growth behaviour of high-strength marine steel based on model of crack tip mechano-electrochemical effect, Journal of Materials Science & Technology 190 (2024) 93–105.
[7] Y. Lu, Y. Li, H. Liu, Y. Zhang, X. Wang, M. Chen, C. Zhao, M.A. Wahab, Learning multi-dimensional sensor relationships for robust marine pipeline leakage non-destructive detection via adaptive graph networks, Eng. Struct. 350 (2026) 121983.
[8] Y. Lu, Y. Zhang, X. Qiu, W. Ren, C. Zhao, M. Chen, Y. Li, S. Bader, H. Liu, Structural health monitoring of offshore pipelines via a novel spatial-topological adaptive graph neural network, Struct. Health Monit. (2026) 14759217261418056.
[9] J. Meriaux, M. Boinet, S. Fouvry, J.C. Lenain, Identification of fretting fatigue crack propagation mechanisms using acoustic emission, Tribol. Int. 43 (2010) 2166–2174.
[10] M. Chai, X. Hou, Z. Zhang, Q. Duan, Identification and prediction of fatigue crack growth under different stress ratios using acoustic emission data, Int. J. Fatigue 160 (2022) 106860.
[11] Y. Zhang, V. Adin, S. Bader, B. Oelmann, Leveraging Acoustic Emission and Machine Learning for Concrete Materials damage Classification on embedded Devices, IEEE Trans. Instrum. Meas. 72 (2023) 1–8.
[12] X. Li, Y. Wang, J. Xing, Y. Wang, Causal graph inference with adaptive dynamic structure learning for mechanism-oriented fault diagnosis in dynamic industrial systems, Reliab. Eng. Syst. Saf. 266 (2026) 111865.
[13] Y. Zhou, M. Liang, X. Yue, Deep residual learning for acoustic emission source localization in a steel-concrete composite slab, Constr. Build. Mater. 411 (2024) 134220.
[14] Y. Lu, M. Chen, X. Qiu, W. Ren, C. Zhao, H. Liu, Offshore Platform Pipeline Leakage Valve Localization using DCEEMDAN and ATSFN, J. Offshore Mech. Arct. Eng. 148 (2025).
[15] Y. Lu, Y. Zhang, M. Chen, Q. Liu, X. Wang, Z. Liu, H. Liu, Acoustic emission-based graph learning for internal valve leakage localisation in offshore pipelines, Nondestructive testing and evaluation (2025) 1–29.
[16] Y. Lu, Z. Zhu, H. Liu, M. Chen, X. Qiu, H. Xu, X. Qu, End-to-End graph neural network framework for precise localization of internal leakage valves in marine pipelines based on Intelligent graphs, Adv. Eng. Inf. 68 (2025) 103716.
[17] X. Li, J. Gu, M. Li, W. Fan, Y. Xiao, W. Lyu, C. Du, Y. Wang, A physically inspired and bounded metric learning framework for fault detection and multi-class classification in mechanical systems, Expert Syst. Appl. 302 (2026) 130393.
[18] X. Li, Z. Gao, J. Zhang, K. Zhu, J. Gu, W. Lyu, Y. Wang, A theoretically Grounded Multichannel Fusion Framework Integrating Dual-Metric Trust and uncertainty Rectification for Intelligent Diagnosis, IEEE/ASME Trans. Mechatron. (2025) 1–10.
[19] H. Shi, S. Zhou, Y. Xie, F. Pan, Z. Fang, X.-Q. Jiang, Enhancing lightweight image super-resolution with hybrid convolution and attention, Pattern Recogn. Lett. 199 (2026) 191–197.
[20] C.-C. Yeung, K.-M. Lam, Contrastive decoupling global and local features for pavement crack detection, Eng. Appl. Artif. Intel. 133 (2024) 108632.
[21] C. Lv, C. Han, J. Chen, D. Cheng, J. Qian, TSD-Depth: using transformers and self-distilling for self-supervised indoor depth estimation, Optik 288 (2023) 171219.
[22] C. Xiang, J. Guo, R. Cao, L. Deng, A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex detection scenarios, Autom. Constr. 152 (2023) 104894.
[23] X. Wang, J. Xu, Q. Yue, X. Liu, Tracing fracture damage evolution and identifying damage patterns in cast steel using advanced acoustic emission analysis, Eng. Fract. Mech. 293 (2023) 109680.
[24] J. Huang, Z. Zhang, B. Zheng, R. Qin, G. Wen, W. Cheng, X. Chen, Acoustic emission technology-based multifractal and unsupervised clustering on crack damage monitoring for low-carbon steel, Measurement 217 (2023) 113042.
[25] Y. Li, H.-L. Minh, M. Cao, X. Qian, M. Abdel Wahab, An integrated surrogate model-driven and improved termite life cycle optimizer for damage identification in dams, Mech. Syst. Sig. Process. 208 (2024) 110986.
[26] H. Chu, W. Wang, L. Deng, Tiny-Crack-net: a multiscale feature fusion network with attention mechanisms for segmentation of tiny cracks, Comput. Aided Civ. Inf. Eng. 37 (2022) 1914–1931.
[27] J. Go, J. Ryu, Spatial Bias for attention-free non-local neural networks, Expert Syst. Appl. 238 (2024) 122053.
[28] N. Hu, D. Zhang, K. Xie, W. Liang, K. Li, A. Zomaya, Multi-graph fusion based graph convolutional networks for traffic prediction, Comput. Commun. 210 (2023) 194–204.
[29] Z. He, M. Jia, L. Wang, UACNet: a universal automatic classification network for microseismic signals regardless of waveform size and sampling rate, Eng. Appl. Artif. Intel. 126 (2023) 107088.
[30] Z. Wang, Z. Leng, Z. Zhang, A weakly-supervised transformer-based hybrid network with multi-attention for pavement crack detection, Constr. Build. Mater. 411 (2024) 134134.
[31] R. Zhu, A. Boukerche, L. Feng, Q. Yang, A trust management-based secure routing protocol with AUV-aided path repairing for Underwater Acoustic Sensor Networks, Ad Hoc Netw. 149 (2023) 103212.
[32] I. Beltagy, M.E. Peters, A.J.A. Cohan, Longformer: The Long-Document Transformer, arXiv preprint arXiv:2004.05150. (2020).
[33] J. Wang, Q. Pan, Z. Liu, I.Y. Lu, S. Chen, Z. Kuang, S. Lin, Reinforcement Learning based Edge Offloading Strategy for Port Infrastructure States monitoring, in: 2024 IEEE/CIC International Conference on Communications in China (ICCC Workshops), 2024, pp. 358–363.
[34] Y. Fan, L. Zhang, P. Li, A Lightweight Model of Underwater Object Detection based on YOLOv8n for an Edge Computing Platform, Journal of Marine Science and Engineering (2024) 697.

[35] Y. Xie, Y. Nie, J. Lundgren, M. Yang, Y. Zhang, Z. Chen, Cervical Spondylosis Diagnosis based on Convolutional Neural Network with X-ray Images, Sensors (2024) 3428.
[36] Z. Han, Y. Li, M. Abdel Wahab, Polynomial chaos expansion-driven Bayesian inference for multi-parameter identification of large-span curved footbridge, Structures 80 (2025) 110035.
[37] Y. Li, M.A. Hariri-Ardebili, T. Deng, Q. Wei, M. Cao, A surrogate-assisted stochastic optimization inversion algorithm: Parameter identification of dams, Adv. Eng. Inf. 55 (2023) 101853.
[38] X. Shang, H. Fang, L. Lu, Z. Zhang, P. Ma, Partial gradient-enhanced multi-output sparse polynomial chaos expansion method for model representation, Reliab. Eng. Syst. Saf. 270 (2026) 112132.
[39] X. Zhang, S. Debroy, Resource management in mobile edge computing: A comprehensive survey, ACM Computing Surveys 55 (13s) (2023) 1–37.
[40] B. Lyu, H. Yuan, L. Lu, Y. Zhang, Resource-Constrained Neural Architecture Search on Edge Devices, IEEE Trans. Network Sci. Eng. 9 (2022) 134–142.
[41] Y. Zhang, Y. Xu, L.S. Martinez-Rau, Q.N.P. Vu, B. Oelmann, S. Bader, On-Device Crack Segmentation for Edge Structural Health monitoring, in: 2025 IEEE Sensors Applications Symposium (SAS), 2025, pp. 1–6.
[42] Q. Wei, L. Shen, B. Kövesdi, L. Dunai, M. Cao, A lightweight stochastic subspace identification-based modal parameters identification method of time-varying structural systems, J. Sound Vib. 570 (2024) 118092.
[43] J. Zhang, M. Zhang, D. Wang, M. Yang, C. Liang, Multi-scale convolutional sparse attention transformer: a lightweight fault diagnosis model for rotating machinery, Neurocomputing 650 (2025) 130934.
[44] W. Wu, N. Zhou, X. Liang, W. Gui, C. Yang, Y. Liu, TCAC-transformer: a fast convolutional transformer with temporal-channel attention for efficient industrial fault diagnosis, Expert Syst. Appl. 297 (2026) 129473.
[45] S. Yan, H. Shao, J. Wang, X. Zheng, B. Liu, LiConvFormer: a lightweight fault diagnosis framework using separable multiscale convolution and broadcast self-attention, Expert Syst. Appl. 237 (2024) 121338.